



Modélisation multi-échelles de la morphologie urbaine à partir de données carroyées de population et de bâti

Johanna Baro

► To cite this version:

Johanna Baro. Modélisation multi-échelles de la morphologie urbaine à partir de données carroyées de population et de bâti. Géographie. Université Paris-Est, 2015. Français. <NNT : 2015PESC1004>. <tel-01263031>

HAL Id: tel-01263031

<https://tel.archives-ouvertes.fr/tel-01263031>

Submitted on 27 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de doctorat

Soutenue le 25 mars 2014

Modélisation multi-échelles de la morphologie urbaine à partir de données carroyées de population et de bâti

par **Johanna Baro**

en vue de l'obtention du titre de docteur de

l' Université Paris-Est dans le cadre de

l'école doctorale n° 532 – MSTIC

Structure de recherche d'accueil : GRETTIA/DEST/LVMT

UNIVERSITÉ PARIS-EST
ÉCOLE DOCTORALE MSTIC
MATHÉMATIQUES ET SCIENCES ET TECHNOLOGIES DE
L'INFORMATION ET DE LA COMMUNICATION

THÈSE

présentée en vue d'obtenir le titre de

Docteur de l'Université Paris Est

Spécialité : SCIENCES ET TECHNOLOGIES DE L'INFORMATION
GÉOGRAPHIQUE

par

Johanna BARO

Modélisation multi-échelles de la
morphologie urbaine à partir de données
carroyées de population et de bâti

VOLUME I

Thèse préparée à l'IFSTTAR,
au sein des laboratoires GRETTIA, DEST et LVMT

Soutenue publiquement le 25 mars 2015 devant un jury composé de :

Pierre FRANKHAUSER	Université de Franche-Comté	(Rapporteur)
Charles BOUVEYRON	Université Paris Descartes	(Rapporteur)
Michel VERLEYSSEN	Université Catholique de Louvain	(Examineur)
Christiane WEBER	Université de Strasbourg	(Examineur)
Patrice AKNIN	IFSTTAR	(Directeur)
Jean-Paul HUBERT	IFSTTAR	(Co-directeur)
Étienne CÔME	IFSTTAR	(Encadrant)
Olivier BONIN	IFSTTAR	(Encadrant)

Remerciements

En premier lieu, je tiens à remercier mon directeur, mon co-directeur ainsi que mes encadrants, Patrice Aknin, Jean-Paul Hubert, Étienne Côme et Olivier Bonin. Je les remercie pour leur disponibilité et le temps qu'ils m'ont consacré, pour leurs précieux conseils, et pour leur relecture attentive de tous mes travaux. Je les remercie de m'avoir laissé avancer librement entre deux disciplines tout au long de mon travail, et d'avoir su m'aider à conserver un équilibre entre les deux. Ce fut un plaisir de travailler avec eux sur ces questions transversales de recherche.

Je remercie également Pierre Frankhauser, Charles Bouveyron, Christiane Weber et Michel Verleysen d'avoir accepté de participer à mon jury de thèse. Je les remercie de l'intérêt porté à mes travaux et des échanges enrichissants qui ont eu lieu lors de ma soutenance. Et plus particulièrement je remercie les rapporteurs Pierre Frankhauser et Charles Bouveyron pour leur lecture attentive de mon manuscrit et pour leurs remarques constructives.

Je poursuivrai en remerciant l'ensemble des personnes que j'ai eu l'occasion de côtoyer pendant mes années de thèse. C'est un merci général qui je l'espère saura m'excuser auprès des personnes qui ne sont pas mentionnées nominativement, mais qui ne sont pas pour autant oubliées.

Je remercie tous les membres du LVMT pour l'accueil qu'ils m'ont fait lors de ma première année. Et plus encore tous les membres du GRETTIA pour l'accueil qu'ils m'ont fait lors de mon stage et de mes dernières années de thèse. Je tiens notamment à remercier les membres permanents avec une pensée particulière pour Latifa, Allou, Laurent, Mustapha, Olivier, Annie et Joëlle avec qui j'ai eu le plaisir d'échanger.

Je n'oublierai bien évidemment pas mes collègues doctorants ou maintenant docteurs, avec qui ce fut un véritable plaisir de passer ces moments de thèse et de détente. Je remercie les anciens Zohra, Inès, Nicolas, Rony, Hani et Guillaume ; ainsi que ceux à qui il reste encore quelques mois ou années de travail Dihya, Matthieu, Moncef, Anne-Sarah, Andry et Josquin. Merci pour toutes ces discussions aussi farfelues qu'intéressantes, pour les encouragements et le soutien. Et bien sûr une spéciale dédicace au « gang des filles » pour leur présence et pour leur amitié qui restera bien au delà de ces années de thèse je l'espère.

Je conclurai en remerciant ma famille et mes amis pour leur soutien direct et indirect pendant ces années. Je les remercie pour leur précieuse aide pour la relecture et pour l'organisation du pot, mais aussi pour leur patience et l'intérêt qu'ils ont su montrer pour mon travail qui ont grandement contribué à me permettre d'en arriver là.

Table des matières

Introduction générale	1
1 Description et conceptualisation des espaces urbains	13
Introduction	14
1.1 Éléments de modélisation urbaine	15
1.1.1 Le centre	15
1.1.2 La distance au centre	18
1.1.3 La sectorisation de l'espace	23
1.1.4 La multipolarisation	30
1.1.5 Les systèmes de villes	35
1.1.6 Discussion : la centralité et les périmètres urbains	44
1.2 Proposition pour un modèle simple de structure urbaine	47
1.2.1 La densité	49
1.2.2 Les tissus urbains	51
1.2.3 Les objectifs de modélisation	52
Conclusion	54
2 Quantifier les espaces urbains : construction et choix des variables d'intérêt	55
Introduction	56
2.1 Les critères de description : quelles variables d'étude ?	57
2.1.1 Une lecture morphologique des objets urbains	57
2.1.2 La construction d'indicateurs sur les grilles régulières	65
2.1.3 Discussion	72
2.2 Le choix d'une nomenclature de tissus urbains	75
2.2.1 Modéliser l'organisation urbaine par une structure radioconcentrique	75
2.2.2 Classification des tissus urbains par règles de décision	79
2.2.3 Discussion	87
Conclusion	88
3 Classification des tissus urbains par transfert d'apprentissage	91
Introduction	92

3.1	État de l'art des méthodes de classification	93
3.1.1	Les approches généralistes	93
3.1.2	Les approches par transfert d'apprentissage	101
3.1.3	Les approches spatiales	104
3.2	Transfert d'apprentissage et champ de Markov	115
3.2.1	Modèles de transfert dans le cadre Markovien	115
3.2.2	Estimation de paramètres	117
3.2.3	Expérimentations sur données simulées	122
3.3	Application : la construction d'une base de tissus urbains	126
3.3.1	Mise en œuvre de la base de données	128
3.3.2	Évaluation de la méthode	135
3.3.3	Évaluation de la base de données des tissus urbains	143
	Conclusion	154
4	Représenter les structures urbaines : des tissus urbains aux mor-	
	photypes	157
	Introduction	158
4.1	Le modèle chorématique	159
4.1.1	La démarche chorématique	159
4.1.2	Chorèmes urbains	162
4.1.3	Exemples de modèles particuliers	167
4.2	Les morphotypes : Enrichir les tissus urbains grâce au raisonnement spatial	170
4.2.1	Le raisonnement spatial	171
4.2.2	Règles d'identification des morphotypes	178
4.2.3	Évaluation des classifications : interpréter les différences	188
	Conclusion	192
5	Typologie des structures urbaines	195
	Introduction	196
5.1	Une représentation quantitative des structures urbaines	196
5.2	Méthodes d'analyse factorielle	205
5.2.1	Les analyses factorielles multiples	206
5.2.2	Extensions de l'AFM aux données mixtes	211
5.3	Identification des tendances spatialement structurantes	215
5.3.1	Les analyses partielles des groupes de variables	216
5.3.2	L'analyse globale des caractéristiques d'organisation	226
5.3.3	Construction de la typologie par CAH	230
5.4	Chorotypes : la spatialisation des enjeux d'aménagement	236
5.4.1	Les modèles en chorotypes	236
5.4.2	Modèles urbains et enjeux d'aménagement	239
	Conclusion	243

Table des matières	v
Conclusion générale	245
Notations	257
Bibliographie	261

Introduction générale

« Un modèle est toujours une simplification de la réalité, ou plus exactement de la vision qu'on a de cette réalité. Cette simplification est faite dans un but opératoire : l'action, la prédiction, ou l'explication. Un modèle est donc un construit et une représentation. » (Brunet, 1980)

La forme urbaine métropolitaine

Les villes ont été dans un passé récent confrontées à des phénomènes d'étalement, de relocalisation des populations et des activités, qui se sont traduits par des changements dans les structures sociales, économiques et morphologiques. Ces transformations sont associées aux dernières révolutions technologiques en matière de transport et de communication qui ont conduit à une dissolution de la distance. La distance n'opère plus comme une contrainte de proximité dans un espace euclidien, mais comme une contrainte temporelle qui bouleverse la physionomie des espaces urbains.

La ville est le support de mouvements rapides qui concernent les déplacements quotidiens, auxquels les habitants consacrent un budget temps de transport d'une heure par jour en moyenne qui semble rester constant entre les villes et les périodes (Zahavi, 1974). Mais elle subit aussi des transformations lentes qui affectent les bâtiments, les activités ou les fonctions. Ces transformations conduisent à des modifications dans l'organisation des villes composées de lieux différenciés de résidence et d'activité dont la mise en relation est assurée par divers types d'infrastructures de transport. Elles conduisent à définir un nouvel objet urbain que nous qualifions ici d'objet urbain métropolitain en renvoyant au type d'organisation décrit par Ascher (2001) ou Roncayolo (2002).

Nous appelons « objet urbain métropolitain » l'ensemble des espaces bâtis caractérisés par une morphologie variable, par des fonctions variables (résidentielles ou d'activités), qui sont reliés par des flux de mobilité quotidienne et polarisés par un noyau

urbain de forte densité. Cette organisation qui a été identifiée dès la fin des années 1950 dans les plus grandes villes nord-américaines (Gottmann, 1961; Webber, 1996), est devenue la forme d'organisation générale au 21ème siècle dans les pays développés. Nous nous intéressons donc à un objet aux limites floues, se développant à la fois par des mouvements centrifuges qui étendent le bâti et intègrent à sa périphérie des villes, bourgs et villages éloignés, et par des mouvements centripètes de densification de l'existant, tout en étant animé par les mouvements pendulaires des déplacements quotidiens. Cet objet possède une ossature relativement stable appuyée sur les masses bâties les plus importantes reliées par un réseau dense de transport et d'interactions fonctionnelles, que nous qualifions ici de **structure urbaine**.

La structure urbaine est étroitement liée à la **forme urbaine** que nous définissons comme la distribution spatiale de l'ensemble des espaces bâtis et des activités dans cette aire aux limites floues. Composée d'une ville-centre, de banlieues, de zones périurbaines, mais aussi de villes secondaires ou satellites, la forme urbaine est complexe et intègre ces espaces liés non plus par une morphologie régulière et compacte mais par des relations fonctionnelles et hiérarchisées. La forme urbaine correspond donc à l'espace urbain ramené à sa dimension physique et fonctionnelle : aux bâtiments et aux infrastructures auxquels on peut rattacher des fonctions qui permettent de caractériser et différencier les lieux.

La question qui nous intéressera ici sera celle du lien entre la forme et la structure. Notre position de départ est que, si la forme est un phénomène géographique observable et représentable par des données, la structure, elle, est un objet émergeant de son substrat qui ne peut être reconstitué que par interprétation théorique et représentable par des modèles. L'étude de la forme est donc un pré-requis pour identifier et analyser des structures urbaines qui sont pour nous des réinterprétations des formes auxquelles on ajoute une dimension relationnelle. L'étude de la structure se base sur la forme, mais aussi sur un travail d'abstraction et un changement d'échelle d'analyse qui permettra d'identifier les éléments stables : les éléments sur lesquels on peut agir pour changer différentes composantes relatives à l'organisation et au fonctionnement des systèmes urbains. Si l'étude de tels éléments se fait souvent par des approches monographiques concernant un espace urbain particulier, l'approche que nous adoptons se veut elle beaucoup plus générique. Notre démarche part de l'observation de la diversité des formes dans l'idée de questionner la diversité des structures.

La ville, sa structure et les modèles

Ces transformations structurelles par croissance ou par absorption des agglomérations périphériques produites depuis les années 70 sont associées à une forte hausse de la consommation d'espace, une fragmentation spatiale et sociale, ainsi qu'une augmentation des déplacements qui s'organisent selon des schémas de plus en plus complexes.

Ces effets aux conséquences négatives d'un point de vue social et environnemental représentent des enjeux actuels d'aménagement qui répondent à la problématique du développement durable introduite par le fameux rapport [Brundtland \(1987\)](#).

En s'intéressant à ces questions, nombreuses sont les études qui se sont penchées sur la façon dont la forme urbaine, prise uniquement dans sa dimension physique associée à une gamme variée de fonctions, conditionnerait les pratiques de déplacement et la consommation énergétique. La densité, en tant que mesure et outil de représentation, constitue alors un des moyens de référence pour qualifier et mesurer l'impact des formes urbaines sur d'autres caractéristiques d'organisation et de fonctionnement. La relation entre ces thématiques aboutit encore à des résultats contradictoires, mais des études de synthèse comme celles de [Stead et Marshall \(2001\)](#); [Ewing et Cervero \(2010\)](#) ont montré que la compréhension de ces liens passe par une lecture des formes urbaines en termes de taille, de forme (géométrique), d'occupation des sols, de relations entre les espaces urbanisés existants et les infrastructures de transport, mais aussi en termes de localisation des nouveaux espaces de développement urbain.

Ces phénomènes posent donc la question de la forme d'organisation optimale qui permettrait de réduire ces divers coûts (distance et nombre de déplacements, temps de trajet et congestion, consommation énergétique, pollution, etc.) ([Tannier, 2009](#)). Cette question demande de décrire finement les formes d'organisation afin de comprendre les conséquences de ces processus de restructuration des espaces urbains. Elle demande aussi de construire des modèles d'organisation pour expliquer, pour comprendre la formation, le fonctionnement et les changements qui s'opèrent dans ces systèmes urbains ([Baumont et Huriot, 1995](#)). Les modèles opèrent comme des représentations permettant de réfléchir sur ces diverses questions d'aménagement. Ils peuvent être conçus comme des formes idéales qui devraient être développées. Mais nous nous intéressons ici aux modèles conçus comme des représentations simplifiées qui mettent l'accent sur les articulations principales entre des formes composant des structures urbaines observées à partir de données.

Dans ce cadre, un modèle de ville compacte et dense a été dans un premier temps opposé un modèle de ville étalée. Ce modèle compact renvoie à une forme caractéristique qui serait celle de la ville européenne ([Newman et Kenworthy 1989](#); [CE, 1990](#)). Il met en avant l'idée que la densité et la compacité engendreraient des distances plus courtes, permettraient un meilleur accès aux transports collectifs et une meilleure mixité sociale et fonctionnelle. Les critiques pointent toutefois sur les limites fonctionnelles de la ville compacte soumise à des problèmes de congestion, d'accès aux aménités environnementales et de coûts élevés pour le foncier et la construction. A l'opposé, le modèle de la ville étalée que l'on pourrait rapprocher d'une organisation caractéristique des villes américaines est lui aussi mis en avant ([Gordon et al., 1989](#)), puisqu'il ne serait pas nécessairement moins favorable car générateur de trajets plus courts et de plus de proximité entre résidences et emplois. Les critiques émises à l'encontre de ce modèle

concernent la dépendance automobile et la consommation d'espace que génère ce type d'organisation (Jenks et al., 1996). À côté de ces deux modèles, d'autres auteurs (Gordon et al., 1991) proposent de se tourner vers des solutions d'organisation multipolaire et hiérarchisée qui pourraient être vues comme un intermédiaire pour concilier principes de développement durable et attentes des individus en termes de localisation et de cadre de vie. Ce modèle qui permettrait de réduire les distances, les temps de trajets et la congestion entraînée par la convergence de tous les flux vers un centre unique, est lui aussi contesté par des résultats empiriques contraires aux hypothèses avancées.

Ce débat entre villes compactes, étalées ou multipolaires est en réalité bien plus complexe que ce que les modèles pourraient laisser penser, puisqu'aucune configuration ne ressemble parfaitement à ces idéaux d'organisations. Nombreuses sont les études empiriques ayant nourri ce débat, qui arrivent à des résultats contradictoires (Crane, 2000; Schwanen et al., 2004; Banister, 2007). Généralement ces travaux se présentent comme des cas spécifiques qui n'étudient qu'une facette de l'organisation. Ils sont souvent observés à une échelle donnée déterminée par les données disponibles. Et ils ne concernent que quelques cas d'étude correspondant généralement aux plus grandes métropoles, ce qui peut poser des problèmes d'interprétation et de transférabilité des résultats trouvés.

Certaines caractéristiques, comme la densité, la mixité fonctionnelle, la multipolarité, l'aménagement local et l'accessibilité en faveur des transports collectifs ou de la marche à pied, sont alors mises en avant pour définir divers modèles de ville durable (Jabareen, 2006). Mais il y a toujours à l'heure actuelle un besoin de connaissance empirique des organisations urbaines quelle que soit leur taille, et leur niveau dans la hiérarchie urbaine. Notre questionnement et notre travail se situent dans le cadre de ce besoin de connaissance qui amènera les clés pour analyser la diversité des configurations existantes, pour comprendre la complexité de ces effets de restructuration des espaces urbains et à terme valider la pertinence de ces caractéristiques.

L'accès à des bases de données de plus en plus fines

L'accès à des bases de données spatialisées de plus en plus précises et exhaustives constitue un atout pour affiner la connaissance des organisations urbaines. Elles offrent le support pour la réalisation d'études statistiques qui permettent de contribuer à faire avancer le débat sur les formes de la ville durable en précisant des modèles d'organisations construits à partir de données détaillées. À ce titre, nous nous intéressons à une tendance récente dans la diffusion des données statistiques qui consiste à s'appuyer sur un maillage régulier à maille carrée, c'est-à-dire un carroyage. L'utilisation des carroyages pour l'analyse de données statistiques est relativement ancienne (Haggett, 1973; Lajoie, 1992), et certains recensements en Europe de Nord utilisent ce type de données depuis les années 1970-1980 (Vaattovaara, 2001; Martin, 2009). Cependant, la

préconisation de l'utilisation de carroyages pour la diffusion de données statistiques au niveau européen par la directive INSPIRE (Annoni, 2003) a donné un nouveau souffle à ce format de données. Dans le cadre de cette directive, l'INSEE diffuse librement depuis 2011 une série de variables socio-économiques sur la population sur des grilles à mailles de 1 kilomètre et 200 mètres de côté couvrant la France entière.

Ce type de maillage apparaît comme une solution pour l'étude de « certains phénomènes qui échappent à toute partition de l'espace, ou tout du moins pour lesquels aucune partition administrative particulière ne s'impose à la raison » (Lajoie, 1992). Cet échantillonnage homogène a montré son intérêt dans diverses études de la morphologie intra-urbaine (Mercier et al., 2008; Hucy, 2010; Dijkstra et Poelman, 2012), de l'extension urbaine modélisée par automate cellulaire (Wu et Martin, 2002), de la modélisation de mesures d'accessibilité (Borzacchiello et al., 2010) ou même de l'évaluation de l'environnement urbain (Brainard et al., 2002).

Ce type de données est très similaire aux données raster, et se prête aux techniques de traitement d'image. Une maille carrée présente l'avantage de constituer l'assise d'un découpage homogène du territoire, correspondant à un échantillonnage spatialement uniforme du phénomène qu'on cherche à mesurer, et neutre du point de vue applicatif. Cette neutralité constitue un atout pour des études comparatives puisque l'homogénéité des unités spatiales donne plus d'assurance pour attribuer des différences de résultats à des différences de spatialisation des phénomènes étudiés et non pas à des différences de signification des unités spatiales.

Les méthodes d'analyses qui peuvent être employées pour étudier des formes d'organisation à détecter sur les grilles régulières doivent toutefois être capables de tenir compte de certaines caractéristiques statistiques propres aux données spatialisées (Miller et Han, 2009; de Smith et al., 2014). Pour proposer une analyse pertinente des données sur grille régulière, il est nécessaire de tenir compte des effets suivants :

- *L'autocorrélation spatiale* : la présence de dépendances statistiques entre les valeurs des variables pour des unités spatiales voisines est l'une des caractéristiques les plus importantes des données spatialisées. Elle est source de complexité dans les traitements statistiques puisque de nombreuses méthodes standard reposent sur des hypothèses d'indépendance. Elle traduit la nécessité de faire appel à des méthodes statistiques spécifiques capables de prendre en compte ces dépendances.
- *L'hétérogénéité spatiale* : l'hétérogénéité spatiale fait référence à la non-stationnarité de divers processus géographiques, qui (par définition) varient selon la localisation. Cette caractéristique implique de porter une attention particulière aux paramètres estimés à un endroit spécifique, pour un jeu de données particulier, qui peuvent subir des effets de dérive par changement de localisation.
- *Les effets d'échelle et de maillage* : tout maillage géographique qui représente le

contenant de mesures statistiques est associé à une unité spatiale de référence qui impose un point de vue sur l'analyse. Dépendant de l'échelle choisie, de caractéristiques géométriques et topologiques, le maillage assure une homogénéité des individus agrégés qui n'a de sens qu'en fonction de la thématique qui a façonné sa construction, et qui va donc façonner son analyse (Openshaw, 1984).

- *La dimensionnalité des données* : la grande dimensionnalité n'est pas une caractéristique propre aux données spatialisées, mais elle est une caractéristique à prendre en compte puisqu'elle peut affecter les analyses statistiques qui deviennent coûteuses en temps de calcul, peuvent subir des effets de la redondance des variables et être soumises à la « malédiction de la dimensionnalité » (Bellman, 1961).

L'apprentissage statistique et le raisonnement spatial : des outils pour la classification de données spatialisées

La démarche d'analyse entreprise repose sur une étude des structures urbaines à partir de données sur grille régulière, où une connaissance *a priori* sert à guider les processus d'exploration des données. Dans ce cadre, différents champs de l'intelligence artificielle ayant trait à l'apprentissage statistique (Hastie et al., 2009), à la représentation de la connaissance et au raisonnement (van Harmelen et al., 2008) seront mobilisés pour proposer des modèles d'organisation urbaine à partir d'une étude de diverses configurations urbaines observées empiriquement.

L'apprentissage statistique permet de réaliser des tâches de traitement et d'analyse de données telles que la classification ou la régression, la sélection ou l'extraction de variables de façon automatique. Ce vaste champ d'étude à l'interface des mathématiques et de l'information a pour objectif d'analyser des données constituant des exemples pour inférer des outils de prédiction ou d'aide à la décision sur de plus grands volumes de données. Dans le cas d'une tâche de classification l'objectif est d'utiliser des exemples tirés d'une population statistique pour construire une fonction mathématique qui puisse prédire la classe de nouvelles réalisations tirées de cette même population en tenant compte de différentes caractéristiques statistiques.

Quelles que soient l'échelle d'étude et l'unité spatiale élémentaire considérées, la classification est la tâche centrale de notre travail puisqu'elle nous permet d'identifier et de classer des ensembles homogènes correspondant à différents types d'objets urbains. D'autres tâches d'analyse de données sont réalisées au cours de ce travail, mais le problème principal de notre étude reste celui de la classification de données.

Différents contextes d'apprentissage peuvent être utilisés pour analyser et classer des données. De manière classique, on oppose apprentissage supervisé et non supervisé.

Dans le premier cas, le principe d'apprentissage repose sur l'utilisation d'un échantillon de données représentant des exemples connus qui fixent les classes à apprendre et déterminent la fonction de classification. Une telle approche présente un intérêt pour dégager des formes d'organisation urbaine comparables, puisqu'elle assure que les différents territoires étudiés auront des descriptions similaires conformément aux prototypes de classes pré-déterminés par des exemples. Dans le deuxième cas, l'apprentissage se base sur des exemples partiellement connus. Il s'agit d'identifier une structure pertinente de regroupement des données, qui doit être apprise sans information sur des étiquettes de classes. Une telle approche présente aussi un intérêt pour classer des types généraux d'organisation dans le but de faire ressortir des similarités et régularités qui ne sont pas nécessairement connues.

Dans le cas du traitement de données spatialisées, l'emploi d'une méthode de classification supervisée est compliqué par l'hétérogénéité spatiale qui génère une non-stationnarité de distribution incompatible avec l'utilisation de ces méthodes. Si les exemples servant à construire la fonction de classification sont tirés selon une distribution différente de celle des nouvelles données à classer, alors le classifieur est invalide. Dans ce cadre, le domaine émergent du transfert d'apprentissage ([Pan et Yang, 2010](#)) trouve un intérêt majeur pour gérer la classification de multiples territoires de façon automatique. Le principe général est d'utiliser la connaissance acquise lors d'un problème d'apprentissage pour améliorer la réalisation d'un autre problème. Dans le cas de problèmes de classification supervisée, le transfert d'apprentissage permet d'adapter un classifieur construit à partir d'exemples associés à un territoire donné pour le rendre utilisable sur d'autres territoires. Il permet de traiter des territoires où des exemples ne sont pas forcément disponibles, et où rien n'assure que les distributions statistiques et spatiales seront les mêmes que sur le premier territoire. Ce cadre de travail représente donc un premier enjeu pour définir les processus d'analyse adaptés au traitement des grilles régulières.

Mais nous verrons que le raisonnement spatial constitue un autre outil indispensable pour compléter l'analyse des données sur grille régulière. Le raisonnement spatial peut servir d'outil d'analyse de données et d'inférence pour proposer des représentations des structures urbaines en intégrant les relations entre les objets urbains. Il offre les moyens de formaliser une représentation complète de l'espace géographique dans un cadre informatique. Un cadre dans lequel il sera possible de raisonner en faisant appel à des descriptions qualitatives permettant de qualifier des objets dans un espace relationnel. Cette approche trouve tout son sens pour des études de configurations spatiales où l'aspect relationnel est primordial pour comprendre le fonctionnement urbain. Elle définit donc un second enjeu pour définir les processus d'analyses des grilles régulières, qui sera considéré conjointement dans ce travail d'analyse des organisations mené à plusieurs échelles d'études.

Positionnement et démarche de la recherche

Le problème auquel nous nous intéressons ici se situe dans le cadre de toutes les thématiques abordées précédemment. L'objectif de ce travail est de proposer une nouvelle représentation des organisations urbaines, élaborée dans une démarche de modélisation assise sur un traitement automatique de données statistiques et géographiques. Il s'agit de fournir une connaissance détaillée des caractéristiques d'organisation les plus marquantes de différents territoires urbains français à partir du traitement de données de densité sur grille régulière.

Cette démarche de modélisation qui implique de combiner des enjeux aussi bien thématiques que méthodologiques, repose sur la mobilisation de plusieurs domaines de compétence. Les différentes contributions de ce travail de thèse sont ainsi articulées autour du traitement de données sur grille régulière, et répondent aux objectifs suivants.

a) Caractériser et comparer des structures urbaines dans l'idée de construire un référentiel d'étude et de connaissance de la ville qui puisse être reconnecté à des contenus relatifs à la mobilité et la morphogenèse.

Nous proposons de nous intéresser à la forme des organisations intra-urbaines observées à partir des traces laissées par les hommes dans le cadre physique de la ville. Ces organisations seront observées dans le périmètre des aires urbaines françaises. Ces périmètres sont considérés comme représentatifs de l'étendue des objets urbains métropolitains puisqu'ils sont des espaces relationnels définis en termes de flux de mobilité quotidienne. Nous entreprenons un travail descriptif et analytique des formes urbaines observées d'après des données produites autour des années 2007-2011. L'objectif est d'en déduire une structure expliquant les conditions de l'accueil de personnes et d'activités ainsi que des interactions et des flux réalisés au sein de l'objet urbain métropolitain (Webber, 1996). Il est donc question d'apporter une connaissance qui puisse être exploitée pour répondre aux questions soulevées par la recherche de forme de ville durable et optimale, une connaissance qui soit un support de compréhension des mouvements inscrits aussi bien dans des temporalités courtes (les déplacements quotidiens des individus), que longues (la morphogenèse du système).

L'enjeu est donc de pouvoir construire des représentations comparables sur l'ensemble des 50 plus grandes aires urbaines françaises, Paris mis à part, à partir de données de densité sur grille régulière. Il s'agit de constituer un référentiel simple et robuste, sous forme d'images classées, qui apporte un premier niveau de description des espaces urbains, qui puisse être enrichi dans le futur de nouvelles thématiques d'informations pour affiner la connaissance des espaces urbains.

Le support neutre des grilles régulières servira de référentiel permettant de s'affranchir de certains problèmes de mesures de densité, liés à l'utilisation de maillages

administratifs non adaptés. Les données de population diffusées par l'INSEE sur des grilles à mailles de 200 mètres de côté croisées avec d'autres données relatives au cadre bâti seront utilisées pour classer des régularités statistiques et identifier des marqueurs de certains types de configurations urbaines.

b) Proposer des méthodes de classification adaptées au traitement de données spatialisées présentant des phénomènes d'autocorrélation spatiale et d'hétérogénéité spatiale

La volonté de mener ce travail exhaustif soulève de nombreux enjeux méthodologiques relatifs au traitement de territoires si variés. Le passage de modèles d'organisation formalisés de manière théorique à des données empiriques est un travail délicat qui doit être entrepris à l'aide de différentes méthodes d'analyse de données. L'exploitation des données sur grille régulière passe notamment par l'utilisation de méthodes de classification adaptées aux caractéristiques des données spatialisées. Mais elle passe aussi par l'utilisation d'une méthode de traitement de données nous permettant de reconstruire des structures urbaines observées afin de les confronter à des modèles d'organisation théoriques.

Vu dans une démarche de modélisation, l'enjeu est ici d'établir le lien permettant de passer d'un domaine conceptuel à un domaine empirique (Livet et al., 2008). La complexité des objets modélisés, autant que l'objectif de fournir une étude exhaustive des organisations urbaines sont des contraintes fortes pour opérer ce passage entre les domaines. Du point de vue du traitement des données l'enjeu est double puisqu'il s'agit de prendre en compte les dépendances statistiques qui génèrent de l'autocorrélation spatiale pour produire des classifications pertinentes des données de densité. Mais il s'agit aussi de prendre en compte la variété des aires urbaines qui doivent être classées. Dans une démarche de classification supervisée, la diversité des jeux de données à classer est une source d'hétérogénéité des distributions statistiques à laquelle il est nécessaire de s'adapter.

Nous proposons donc de construire des bases de données décrivant les formes bâties à différents niveaux d'organisation afin d'interpréter les caractéristiques d'organisation les plus marquantes. Le traitement des données sur grille régulière passera alors par trois phases d'exploitation aux échelles emboîtées. Nous nous efforcerons de faire correspondre les échelles de trois types d'unités spatiales à trois types d'objets d'urbains :

1. *Pixel* : Nous proposons un algorithme de classification utilisant le formalisme des champs de Markov discrets cachés pour gérer les dépendances spatiales, qui sera appliqué dans un cadre de transfert d'apprentissage pour gérer l'hétérogénéité spatiale. L'application de cet algorithme sur des données traitées à l'échelle du pixel nous permettra de représenter et d'analyser des territoires bâtis spatialisés sous la forme de classifications en tissus urbains.
2. *Région d'image* : Nous proposons ensuite d'utiliser la chorématique et les ou-

tils du raisonnement spatial qualitatif pour opérer un enrichissement des données classées par l'algorithme précédent. En se servant de ces outils d'analyse d'image, nous passons à une représentation déterritorialisée sous la forme d'une classification en morphotypes représentant les objets stables composant la structure urbaine. A cette échelle d'analyse les relations entre objets observées dans un cadre topologique définissent un cadre spatial relatif, utilisé comme base de représentation et de comparaison des structures urbaines.

3. *Image* : L'exploitation de ces classifications en morphotypes sera finalement assurée à l'aide de méthodes d'analyse factorielle de données mixtes et de classification non supervisée. Ces méthodes sont utilisées pour analyser les données traitées cette fois-ci à l'échelle des images sous la forme d'un espace de mesures quantitatives résumant différentes caractéristiques d'organisation. L'analyse de ces caractéristiques est utilisée pour identifier les caractéristiques les plus marquantes des organisations de ces 50 aires urbaines et proposer des modèles chorématiques résumant les différents types de structures urbaines observés à partir des données.

Organisation de la thèse

Plusieurs cheminements sont possibles pour aboutir à la représentation d'un espace et à la constitution d'un modèle. Haggett (1973) identifie trois phases de modélisation qui se nourrissent mutuellement (l'abstraction, l'expérimentation et l'interprétation), que nous retenons pour résumer le processus global de modélisation et présenter l'organisation du manuscrit (figure 1). Ces trois phases sont présentées dans le manuscrit qui s'organise en cinq chapitres.

Le chapitre 1, celui de la « phase d'abstraction », définit l'objet que nous souhaitons analyser et construire. Il pose le cadre de modélisation théorique qui sera utilisé pour guider tous les travaux de mise en forme et de classification des données empiriques. Nous discuterons de l'objet qu'est la structure urbaine, et nous donnerons un état de l'art détaillé des modèles de structure urbaine à l'échelle meso-géographique et macro-géographique. Cela nous mènera à définir le modèle d'organisation que nous souhaitons observer : un modèle d'organisation morphologique interprétant des caractéristiques de densité et de composition en tissus urbains.

Le chapitre 2 poursuit cette phase de modélisation par la question des variables à utiliser pour obtenir une version quantifiée du modèle d'organisation défini de manière conceptuelle. Le choix des variables a une grande influence sur la construction et la définition du modèle puisqu'il conditionne ce qu'il est possible de représenter à partir des données disponibles. Nous y introduirons les données de population de l'INSEE, ainsi que les données sur le cadre bâti tirées de la BD Topo de l'IGN. Puis nous présenterons la série d'indicateurs utilisés dans les traitements mathématiques et la raison

de leur choix. Enfin nous proposons une nomenclature des tissus urbains, appuyée par la réalisation de travaux préparatoires de classification.

Le chapitre 3 introduit la « phase d'expérimentation » vue dans notre cas comme une phase de mise en forme et de classification des données. Il permettra de présenter l'algorithme de classification par champ de Markov caché appliqué dans un contexte de transfert d'apprentissage que nous proposons. Les travaux de classification produits par cet algorithme à la première échelle d'étude du pixel seront présentés ici. Nous détaillerons les résultats des classifications en tissus urbains réalisées sur l'ensemble des 50 aires urbaines étudiées.

Le chapitre 4 poursuit ce travail en proposant une représentation des structures urbaines, au sens de la structure du bâti urbain, observées par un enrichissement sémantique des régions d'image produites par les précédents travaux de classification. En raisonnant à l'échelle des régions labellisées, nous verrons qu'il est possible de définir des règles d'analyse systématique des classifications pour identifier des catégories de morphotypes représentant les éléments stables des organisations représentées par notre modèle théorique. Pour cela nous articulerons une démarche entre chorématique et raisonnement spatial qualitatif pour automatiser la production de ces modèles d'organisation observés sur les grilles régulières.

Enfin le chapitre 5 correspondra à la « phase d'interprétation » de ce modèle d'organisation. Il permettra de tirer des conclusions relatives aux caractéristiques les plus marquantes de ces organisations observées ici dans un espace de caractéristiques quantitatives. Nous proposons dans ce chapitre d'utiliser des méthodes d'analyse factorielle et de classification pour dégager les tendances d'organisation les plus marquantes et obtenir une typologie d'aires urbaines permettant de fournir de nouvelles classes de modèles intermédiaires sous forme de chorotypes.

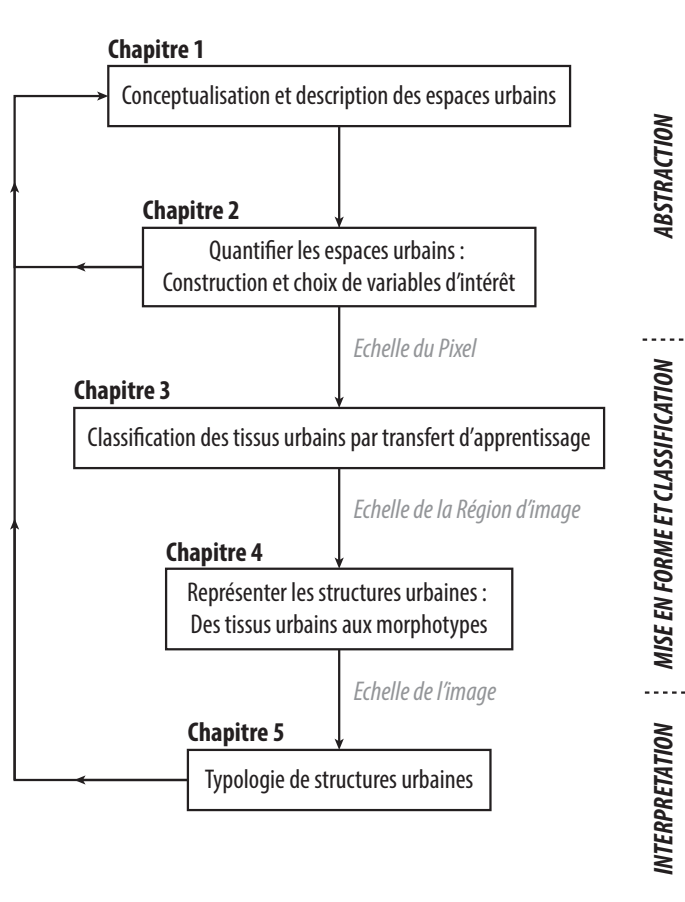


FIGURE 1 – Structure de la thèse

Chapitre 1

Description et conceptualisation des espaces urbains

Sommaire

Introduction	14
1.1 Éléments de modélisation urbaine	15
1.1.1 Le centre	15
1.1.2 La distance au centre	18
1.1.3 La sectorisation de l'espace	23
1.1.4 La multipolarisation	30
1.1.5 Les systèmes de villes	35
1.1.6 Discussion : la centralité et les périmètres urbains	44
1.2 Proposition pour un modèle simple de structure urbaine	47
1.2.1 La densité	49
1.2.2 Les tissus urbains	51
1.2.3 Les objectifs de modélisation	52
Conclusion	54

Introduction

L'objet d'étude est ici la structure des formes urbaines métropolitaines que nous cherchons à approcher à partir des modèles d'organisation urbaine de la littérature en géographie ou économie spatiale. La difficulté à définir une structure urbaine tient à la complexité et à la diversité des espaces urbains et au fait qu'il est possible d'y mettre plusieurs contenus : un contenu physique et paysager relatif au cadre bâti et aux infrastructures, et un contenu relatif au cadre social, aux individus et activités économiques qui sont rattachés et spatialisés dans ce cadre bâti.

La structure est un outil pour comprendre l'organisation des formes urbaines mais elle n'est pas directement observable en leur sein. Elle demande une interprétation théorique des différents modes d'organisation de leur substrat, qui est pour nous le cadre physique et fonctionnel qui abrite individus et activités. Les formes physiques fournissent un référentiel concret de la ville, qui permet de visualiser facilement à partir de données ces phénomènes d'organisation, d'en extraire les régularités et les éléments à forte inertie. Elles permettent alors d'appréhender des structures d'organisation plus complexes et variées : « Évocatrices de leur genèse, les formes urbaines, rehaussées par l'image, deviennent les témoins de l'action des sociétés qui les créent, les utilisent et les interprètent. » (Weber, 1995, p. 85)

Derrière l'objet visé se cache donc la volonté d'observer et d'analyser des données urbaines relatives à ces formes physiques afin de faire ressortir par des méthodes mathématiques des éléments de représentation de plus en plus abstraits, qui entrent dans une démarche de modélisation. La suite de ce chapitre explique cette démarche de modélisation/abstraction à partir de données. Il a pour objectif de poser le cadre théorique qui sera utilisé afin de proposer des éléments de modélisation simples guidant la construction et l'analyse de classification des données sur une grille régulière.

Pour cela nous présenterons d'abord un état de l'art des modèles de villes ou de régions urbaines qui serviront de support de réflexion pour la construction de nos éléments de modélisation. Nous verrons que ces différents modèles théoriques contiennent beaucoup d'éléments pertinents, mais que tous ne pourront pas être utilisés, notamment à cause des données disponibles. La deuxième partie du chapitre sera donc consacrée à exposer les éléments de modélisation que nous retenons pour construire une représentation simple des structures urbaines. Nous introduirons les notions qui serviront à définir la manière dont nous souhaitons construire ces structures urbaines. Nous verrons qu'une représentation à grande échelle de la morphologie concrète des espaces urbains, sous l'angle de la densité et de la composition selon sdes tissus urbains amènera à définir un support de compréhension et de différenciation des structures urbaines.

1.1 Éléments de modélisation urbaine

La recherche de structures, de régularités, appelle à se tourner vers un autre concept, celui de modèle. Le modèle fournit un moyen de représenter ces régularités, de les expliquer, de les retransmettre. Il permet de simplifier la réalité trop complexe pour être appréhendée de façon exacte, et de mettre de la généralité pour figurer que les espaces urbains, aussi divers et particuliers qu'ils soient, comportent des similitudes de développement qui les rendent comparables.

Nous nous attachons à présenter dans cette section un état de l'art des modèles permettant de représenter des organisations urbaines observées de l'échelle meso-géographique détaillant les organisations intra-urbaines, à l'échelle macro-géographique détaillant des régions urbaines ou systèmes de villes. Cet état de l'art se compose de modèles anciens, de leurs actualisations, et de modèles plus récents. Tous ces modèles qui ont été construits pour répondre à des objectifs particuliers, à partir de données particulières, présentent un intérêt pour comprendre au moins partiellement les processus actuels de structuration des espaces urbains français.

Cette présentation de l'état de l'art nous permet d'analyser le lien entre chaque modèle et sa base empirique, qui n'est identique pour aucun des modèles et qui ne sera pas identique à la nôtre. C'est aussi et surtout un moyen de mettre en avant les éléments stables qu'il serait pertinent d'intégrer pour construire une représentation « actualisée » des structures urbaines.

1.1.1 Le centre

« Il n'y a ville que s'il y a un noyau susceptible d'unifier, de dominer, d'organiser la périphérie. » (Beaujeu-Garnier, 1995, p. 15)

Posé comme le lieu moteur de l'urbain, le centre est l'élément organisateur qui distinguera par sa présence les organisations urbaines. Le concept de centre, autant que l'objet morphologique qu'il représente constituent des éléments majeurs de la modélisation urbaine et plus généralement de l'analyse spatiale. Nous ne nous intéressons pas aux raisons de la création de ce centre organisateur, mais à ses caractéristiques concrètes en tant qu'élément de la structure, et à son rôle en tant qu'élément organisant le reste de la structure. Les définitions sont diverses, tant le concept trouve une place importante dans toutes les disciplines attachées à étudier le champ de l'urbain (géographie, économie, sociologie, histoire, etc.). Nous donnerons ici les éléments de définitions repris dans les différentes modélisations urbaines de la géographie et des sciences régionales.

1.1.1.1 Éléments de définitions

Élément d'une organisation spatiale, le centre est déterminé par des caractéristiques symboliques associées à la présence de certaines fonctions de pouvoir et de formes de prestige liées à l'histoire des lieux. Certaines de ces caractéristiques sont observables, mais toutes semblent difficilement mesurables en raison de la place laissée à l'interprétation de leur symbolique. Un centre est un lieu qui bénéficie de toutes les caractéristiques de la centralité et comme le soulignent [Huriot et Perreur \(1997\)](#), deux dimensions donnent les éléments fondamentaux de définition d'un centre :

- *une dimension positionnelle* (et géométrique) : un centre est un lieu qui minimise la distance à tous les autres lieux dans un espace limité ; le choix de la mesure de distance représentant alors un élément essentiel pour déterminer cette position. Une des valeurs fondamentales pour qualifier un lieu de central serait donc sa très bonne accessibilité.
- *une dimension attributaire* : un centre est premièrement un lieu de concentration de la population. Mais il est surtout un lieu de concentration du pouvoir politique et administratif, des fonctions économiques, de l'offre de biens et services, de la culture, de la connaissance, de l'information et de l'innovation. Aujourd'hui tous ces attributs ne sont plus nécessairement présents simultanément dans les lieux qualifiés de centraux. Une nouvelle subtilité apparaît dans le concept de centre qui sera discutée plus tard, à savoir que les centres peuvent être spécialisés et donc multipliés.

Ces dimensions complémentaires permettent de retrouver la caractéristique principale d'un centre qui apparaît comme le lieu privilégié des contacts, des interactions et des échanges ([Bourdeau-Lepage et al., 2009](#)). Dans une organisation urbaine, le centre est un élément essentiel puisqu'il représente le principal lieu de réception de flux, qu'ils soient humains ou qu'ils concernent les activités économiques, la connaissance et les informations.

L'encadré 1.1 présente les deux principales déclinaisons du concept de centre en notions reprises dans les modélisations de la ville. Deux notions de centre-ville plutôt qu'une sont nécessaires pour caractériser les centres observés : une pour les modèles de villes européennes (le centre historique) et une pour les modèles de villes américaines (le CBD). Ces notions ne sont pas exclusives à une de ces localisations, et d'autres notions pourraient encore être ajoutées pour décrire les situations observables dans les autres parties du monde. Ces notions traduisent que la polysémie du concept de centre est liée à la variété des configurations observables, elle même liée à la localisation et à l'histoire des villes. De plus un centre n'existe qu'à une échelle donnée et que par rapport à une fonction donnée ([Huriot et Perreur, 1997](#)). Les attributs de la centralité ne seront pas les mêmes selon l'échelle considérée, qui peut faire varier la me-

sure de distance considérée pour définir l'accessibilité. De même qu'elle fera varier les types et les concentrations de fonctions à prendre en compte comme attribut générant des interactions et un rayonnement polarisateur. Dans toute modélisation le choix de l'échelle est crucial puisqu'il influencera la représentation et fera observer, par exemple un unique centre multifonctionnel ou plusieurs centres monofonctionnels à une plus grande échelle géographique. Le centre reste toutefois un élément structurel typique et réel bien qu'indirectement observable de façon univoque.

Encadré 1.1 (*Les notions de « centre »*)

Centre historique : « Quartier ancien d'une agglomération, correspondant généralement à son étendue antérieurement à la révolution industrielle et aux secteurs où étaient regroupés les principaux édifices commerciaux, religieux et administratifs » (Gauthiez, 2003). Notion d'architecture et des études d'art et d'histoire, la définition du centre historique contient une forte dimension morphologique puisque ces espaces sont avant tout reconnaissables à leur voirie étroite et à leur parcellaire morcelé. Cependant cette notion peut contenir une certaine ambiguïté liée au contexte local et historique de chaque ville qui pourra donner des objets aux caractéristiques très variables (Merlin et Choay, 1988).

CBD (Central Business District) : Quartier des affaires d'une ville, aussi appelé *Downtown*, le CBD se caractérise autant par ses fonctions économiques de haut niveau dédiées au tertiaire que par sa morphologie marquée par une haute concentration de bâtiments de grande hauteur. Il est l'espace central où la rente foncière est la plus élevée et donc où seules les activités les plus rentables, prestigieuses et intensives en consommation d'espace peuvent se maintenir (Merlin et Choay, 1988). Il représente le centre des villes dans les modèles développés pour décrire les villes américaines au début du siècle dernier (cf. sections 1.1.2 et 1.1.3).

1.1.1.2 Le modèle centre-périphérie

Laissé volontairement de côté jusque là, le concept de périphérie ne peut être ignoré tant il est indissociable de celui de centre. La périphérie est définie par complémentarité au centre. L'un n'existe pas sans l'autre puisqu'une des caractéristiques indispensables pour définir un centre est la présence de relations qui existent entre le lieu qualifié de central et ceux qui seront qualifiés de périphériques. Deux visions peuvent être associées au modèle centre-périphérie, selon l'approche considérée pour qualifier l'organisation de l'espace. Une première vision, détaillée dans la suite, renvoie à des rapports de complémentarité spatiale produits par la distance. Le modèle centre-périphérie renvoie alors à un modèle de gestion de la distance qui se traduit spatialement par des organisations urbaines marquées par des gradients décroissants qui peuvent caractériser des facettes variées de l'organisation.

Une seconde vision renvoie à un modèle d'organisation caractérisé par des rapports de domination de l'espace central sur l'espace périphérique. La position avantageée du centre, lieu le plus accessible et qui concentre le plus de valeur (humaine, richesse, information, etc.) en fait un lieu privilégié d'interaction qui se maintient par l'entretien de flux et de relations asymétriques avec son espace environnant : la périphérie. Souvent employé pour représenter les configurations spatiales aux échelles mondiale (l'économie-monde de Braudel (1979)) ou régionale (la traditionnelle opposition ville/campagne largement documentée en géographie (Beaujeu-Garnier, 1995, Partie 5)), ce modèle tire sa force de son applicabilité à toutes les échelles spatiales. Reynaud (1981) le montre en se servant du modèle pour décrire les situations de domination des centres des villes en termes d'équipements par rapport aux quartiers périphériques, le poids des métropoles à l'échelle des espaces régionaux, ou encore les relations asymétriques entre pays développés et pays en développement.

1.1.2 La distance au centre

« La ville est un mode d'organisation visant à gérer la distance, ou encore le résultat des compromis réalisés par chaque agent entre son désir d'accessibilité à un lieu pour bénéficier de ses ressources, services et aménités, et l'effort nécessaire pour y parvenir, en termes financier, temporel, matériel ou psychologique. » (Baudelle et Ducom, 2009, P. 86)

Fondamentale dans la modélisation urbaine, la distance intervient comme une mesure de l'espace à franchir pour aller d'un lieu à un autre, mais aussi comme vecteur de discontinuité générateur de lieux différenciés. Elle est fondamentale en géographie puisqu'elle permet à la fois de définir et de situer les lieux entre eux, en tant qu'objets localisés dans un même repère géographique. Intégrée au modèle centre-périphérie présenté dans la section précédente, la distance mesurée par rapport à un centre fournit une clé de lecture complémentaire. Elle révèle la diversité des configurations observées via l'ensemble des transitions possibles pour passer du centre à la périphérie.

1.1.2.1 Les bases historiques du modèle monocentrique

Von Thünen Repris par les économistes, autant que par les géographes et les sociologues, le modèle de localisation agricole de von Thünen (1826) est utilisé dans de nombreux travaux de modélisation urbaine. Ce modèle de localisation, appliqué au milieu urbain, constitue la base de la classe des modèles de la ville monocentrique. Aussi appelé modèle de « l'État isolé » de Von Thünen, ce modèle non mathématique pose les bases de la future théorie de la rente foncière. Le modèle est basé sur les hypothèses suivantes :

- L'espace est une plaine homogène, que ce soit au point de vue de la qualité du

sol, de la topographie ou du climat.

- Au centre de cette plaine homogène se trouve un unique marché d'échange de toutes les productions agricoles de la plaine : la ville.
- Cet espace constitue un « État isolé », à savoir qu'il n'a aucune relation avec l'extérieur. Les seules relations qui existent ont lieu entre la ville et son hinterland.
- Les coûts de transport sont homogènes dans toutes les directions : il n'y a pas de routes.

Dans l'idée que les producteurs agricoles cherchent à maximiser leur profit, Von Thünen explique que l'organisation de l'hinterland se fait en fonction de la distance au marché et du type de produits. L'espace s'organise en cercles concentriques de productions (figure 1.1), puisqu'il n'est rentable de produire certains types de biens qu'à une certaine distance du marché à cause de la rente et des coûts de transport. Ainsi les biens rapidement périssables et plus intensifs en production possèdent des coûts de transport très élevés mais génèrent plus de profit. Ils sont donc localisés à proximité du centre. Les productions plus extensives, générant un profit un peu moins élevé mais ayant aussi un coût de transport moins élevé (bois de chauffage, céréales, élevage) sont localisées plus loin du centre.

Ce modèle pose les bases pour comprendre l'organisation de l'espace urbain polarisé par un centre qui va orienter les processus de localisation des entreprises aussi bien que des individus. Ce centre, lieu de convergence des personnes et des biens, polarise l'espace de façon à créer des gradients décroissants de densité d'activités, des personnes ou encore de la rente foncière.

Burgess Un siècle plus tard, l'organisation spatiale en cercles concentriques de Von Thünen se retrouve dans le modèle du sociologue **Burgess (1925)**, de l'école de sociologie et de géographie de Chicago. Le modèle ne concerne plus ici les localisations agricoles, mais l'utilisation des sols par des individus caractérisés par leur statut culturel, ethnique et socio-économique. Selon une démarche empirique, Burgess développe initialement sur la ville de Chicago un modèle qui décrit un processus de ségrégation spatiale. Cette ségrégation a lieu dans le but de faciliter les interactions sociales et économiques s'appuyant sur des solidarités culturelles. Ce modèle représente ainsi un processus d'ascension sociale qui se traduit par une migration vers la périphérie.

Les auréoles de population s'organisent autour du CBD, surnommé *Loop* à Chicago (figure 1.2), regroupant les entreprises et vers lequel les voies de communication convergent. Au-delà de ce centre, Burgess définit un premier anneau appelé « zone de transition » qui correspond à une zone très dense d'habitat pauvre et qui regroupe les migrants récents et certaines minorités ethniques. Les anneaux suivants correspondent à d'autres zones résidentielles, d'abord celles des classes moyennes constituées des im-

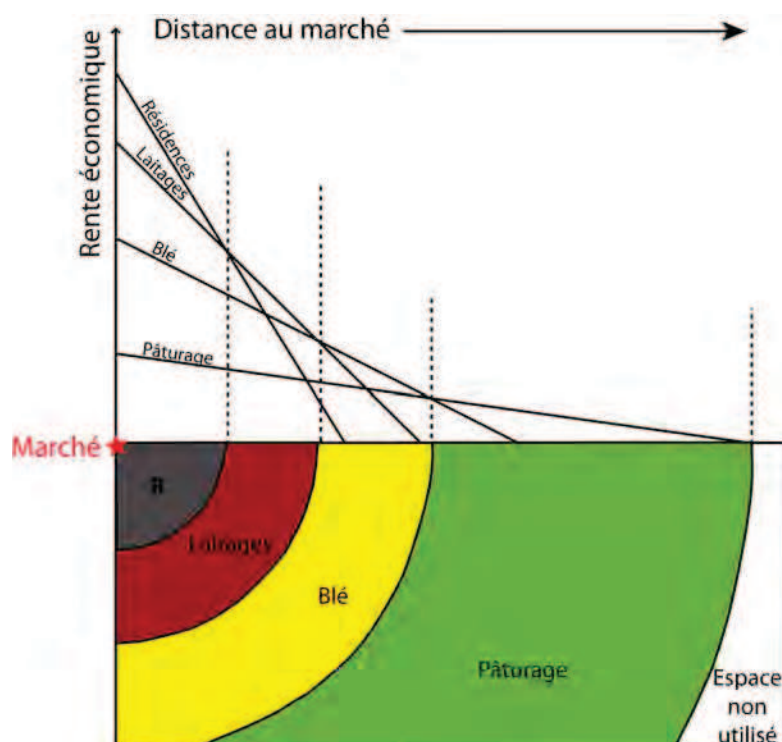


FIGURE 1.1 – Modèle de Von Thünen (base graphique tirée du Oxford Dictionary of Geography)

migrants de deuxième génération, puis celles des classes plus aisées qui se trouvent à la périphérie de la ville. Enfin, au-delà de la ville se trouve l'espace résidentiel des *commuters* qui peuvent se permettre de faire les navettes quotidiennes entre leur lieu de résidence qui n'est plus situé dans l'espace urbain et leur lieu de travail, c'est-à-dire le CBD.

Ce modèle reflète plus particulièrement l'organisation des villes américaines. Le processus de développement urbain se fait par « abandon » des logements les plus anciens et dégradés, c'est-à-dire les logements centraux, laissés aux classes les moins favorisées. Les classes les plus favorisées, elles, peuvent se permettre de se déplacer vers la périphérie pour reconstruire de nouveaux logements.

Même au siècle dernier, plusieurs critiques ont été émises à l'encontre de ce modèle, la première étant que l'organisation régulière en anneaux concentriques ne reflétait pas la réalité des villes américaines. De cette critique découle la seconde qui est que le modèle ne prend pas en compte les voies de communication, les établissements industriels, et notamment l'industrie lourde, qui déjà dans les années 1920 déformaient les anneaux et laissaient des secteurs vides de population. Cette critique amènera Hoyt, chercheur issu de la même École de Chicago, à repenser le modèle pour aboutir au modèle sectoriel, présenté dans la section 1.1.3.1.

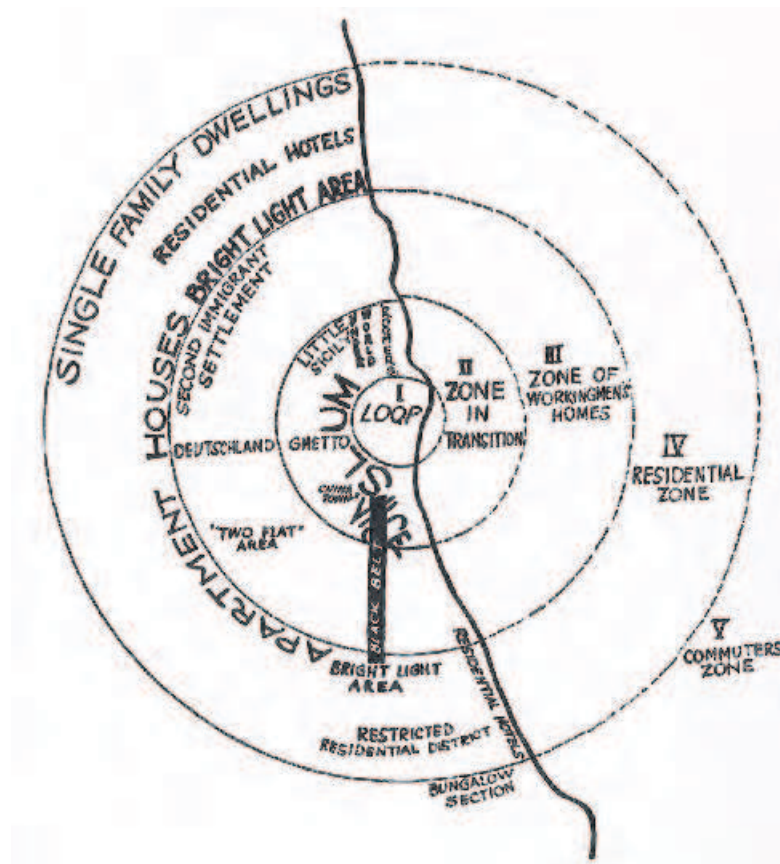


FIGURE 1.2 – Modèle auréolaire de Burgess sur la ville de Chicago, Source : (Burgess, 1925, p. 55)

1.1.2.2 Les organisations urbaines par gradient décroissant

Au delà de sa capacité à décrire une forme d'organisation spatiale alliant un centre et une ou plusieurs couronnes périphériques, l'intérêt du modèle d'organisation concentrique réside dans sa capacité à formuler l'hypothèse encore actuelle d'organisation marquée par des gradients décroissant du centre vers la périphérie. Comme le souligne Racine (1971), même si cette structure concentrique est discutable pour expliquer la localisation des individus caractérisés par des variables économiques ou ethniques, elle est beaucoup plus corrélée à d'autres variables telles que la densité de population ou de construction, de rente foncière, ou encore bien d'autres phénomènes urbains.

Dans ce cadre plusieurs économistes tels qu'Isard (1956), Alonso (1964), Mills (1967) et Muth (1969) se sont interrogés sur les lois régissant l'occupation des sols et les processus explicatifs des systèmes urbains pour aboutir à la théorie de la rente foncière. Cette théorie fournit une explication mathématique des choix résidentiels dans l'espace urbain, comme le résultat d'un compromis effectué par les individus entre l'accessibilité au centre et le loyer à payer pour un terrain de taille donnée. La théorie suppose que

le centre préexistant contient tous les emplois, et que tous les individus résidant dans l'espace urbain subissent un coût de transport pour accéder au centre, sachant que la congestion n'est pas prise en compte. En plus de ces deux hypothèses de base, la théorie repose sur une série d'hypothèses qui traduisent le fait que les loyers doivent être moins chers à la périphérie et les logements doivent être plus grands qu'au centre pour que deux individus restent indifférents à la localisation en conservant la même utilité (déterminée par un ensemble de biens et une surface de logement). De nombreux travaux, parmi lesquels on peut citer l'article d'[Anas et al. \(1998\)](#), ou le livre de [Fujita et Thisse \(2003\)](#) fournissent plus de détails sur la formalisation économique et mathématique de cette théorie.

D'autre part, cette hypothèse d'espaces urbains organisés par des gradients décroissants est utilisée depuis les travaux précurseurs de [Clark \(1951\)](#) sur les densités de population, pour observer la configuration spatiale des agglomérations ou des aires urbaines et les comparer entre elles. Dans ces premiers travaux, le modèle exprime que la densité D_x à une distance x du CBD est une fonction exponentiellement décroissante reliée à la densité au centre D_0 par l'équation :

$$D_x = D_0 \exp(-\gamma x) \quad (1.1)$$

Le paramètre γ correspond au gradient de densité ($\gamma = -D'_x/D_x$) qui exprime la vitesse à laquelle la densité de population décroît avec la distance.

De nombreux auteurs ont depuis contribué à vérifier la conformité des aires urbaines à cette forme de modèle monocentrique en adaptant la fonction utilisée pour représenter le phénomène étudié. On peut notamment citer l'étude des agglomérations américaines d'[Anas et Kim \(1992\)](#) reprenant une fonction exponentielle négative ; l'étude de Séoul de [Batty et Kim \(1992\)](#) qui intègre une approche fractale et se base sur une fonction puissance négative ; l'étude sur Bristol de [Longley et Mesev \(2002\)](#) qui pose la question du référentiel de données utilisé ; l'étude comparative de grandes métropoles du monde de [Bertaud \(2004\)](#) ou encore l'étude européenne des surfaces bâties de [Guérois et Pumain \(2008\)](#) qui se base sur une double régression linéaire.

L'utilisation de ce type de modélisation permet d'appréhender les structures urbaines en termes de dispersion de la population et de tendance à l'étalement. Les comparaisons temporelles de ces gradients, que ce soit en termes de densité de population pour les villes américaines ([Anas et al., 1998](#)) ou en termes de densité de surface bâtie pour les villes européennes ([Guérois et Pumain, 2008](#)), montrent cette tendance à l'étalement extensif. Dans les deux cas, les gradients ont tendance à diminuer en fonction du temps. Cette diminution se traduit par des structures plus étendues et marquées par une organisation plus dispersée. [Anas et al. \(1998\)](#) justifient cette tendance à la dispersion par une hausse des revenus des résidents liée à une baisse des coûts de transport.

Mais comme le montrent [Guérois et Pumain \(2008\)](#), le profil de densité en fonction de la distance au centre révèle, au moins pour les aires urbaines européennes, qu'il existe aussi une tendance à la densification des agglomérations. Les deux tendances coexistent : la tendance à l'étalement concerne la partie discontinue des aires urbaines ; tandis que la tendance à la densification concerne la partie morphologiquement continue de l'agglomération. Ce phénomène de densification est maximal dans ce qui correspond à la première couronne de banlieue entourant le centre historique.

Dans une démarche d'étude visant la compréhension des structures urbaines, le modèle monocentrique constitue donc une représentation incontournable puisqu'il permet de saisir les processus de morphogénèse qui orientent le développement urbain. A l'heure actuelle, compte tenu de la complexité urbaine grandissante, un grand nombre d'agglomérations ne se conforment plus exactement à ce modèle. Mais il n'en reste pas moins la base pour capturer une version simplifiée de l'organisation des espaces urbains, et pour développer de nouveaux modèles intégrant plusieurs centres polarisateurs.

1.1.3 La sectorisation de l'espace

Le modèle monocentrique se fonde principalement sur une interprétation économique des processus de localisation qui se formalisent par le pouvoir attractif et isotrope d'un centre dans un espace vide à deux dimensions réduit à sa plus simple expression. Mais en observant la réalité de la structure interne des agglomérations, il apparaît que la modélisation en anneaux concentriques n'est pas adaptée puisque que de nombreux phénomènes urbains se manifestent par une spécialisation en quartiers de forme non auréolaire. La sectorisation de l'espace apparaît alors comme un élément de modélisation urbaine à part entière qui se retrouve de manière plus ou moins explicite dans beaucoup de modèles urbains.

Les modèles monocentriques les plus simples contiennent eux-même une forme de sectorisation puisqu'ils contiennent deux types d'espaces, le CBD et la zone résidentielle, dont les fonctions distinctes et disjointes traduisent déjà une forme de sectorisation de l'espace. Au-delà de cette simple sectorisation, l'économie spatiale introduit avec les externalités un degré supplémentaire de sectorisation par la spécialisation des quartiers industriels. Les externalités, avantages gratuits dont bénéficient les entreprises à proximité les unes des autres, justifient à la base la concentration des entreprises dans le CBD. Mais ces dernières conduisent en plus à l'apparition de secteurs plus spécialisés dans la ville. Elles conduisent à la spécialisation en quartiers industriels qui ne cohabitent pas toujours à proximité, puisque deux industries très différentes ne s'apportent que peu d'externalités mais subiront les mêmes déséconomies si elles se localisent en un même endroit ([Fujita et al., 1999](#)). Les industries lourdes, par exemple, ne se localisent pas sur les mêmes sites que les activités de commerce et de service, à cause de contraintes de site et d'accessibilité différentes, et aussi des externalités négatives que

génèrent les industries lourdes.

La sectorisation de l'espace ne concerne pas uniquement les activités, elle concerne aussi les personnes. Comme le note [Bailly \(1973\)](#), on peut lire une sectorisation sociale même dans le modèle de Burgess qui est censé faire figurer une homogénéité auréolaire. [Burgess \(1925\)](#) décrit la composition sociale pour chaque anneau, qui se divise en quartiers regroupant des immigrants de même origine. Cette ségrégation, qui est spatialisée dans le modèle particulier de la ville de Chicago (figure 1.2), crée un processus de différenciation sociale et morphologique puisque chaque quartier est associé à un type de population et même à un type de construction.

En complément de ces exemples de processus de sectorisation, deux catégories de modèles faisant de la sectorisation une composante principale de l'organisation des espaces urbains sont présentés dans les sections suivantes. Les modèles historiques de l'école de Chicago sont présentés dans un premier temps, suivis de l'approche structuraliste de l'occupation de l'espace.

1.1.3.1 Les modèles de l'école de Chicago (Hoyt, Harris et Ullman)

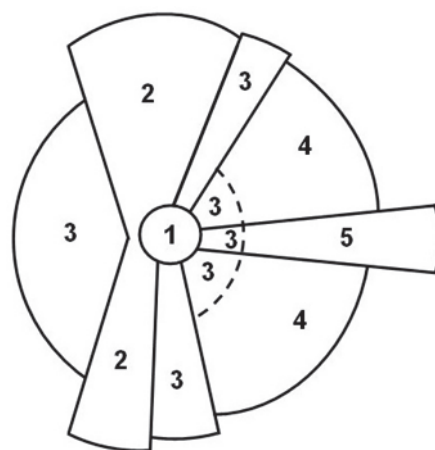
Hoyt Le modèle fondamental de la ville sectorisée est celui de l'économiste américain Homer Hoyt. Son modèle s'inscrit dans le courant de pensée de l'école de Chicago et correspond à une évolution du modèle de Burgess intégrant une réponse à la principale critique faite à son encontre, à savoir l'absence de prise en compte des réseaux de transports.

Selon [Hoyt \(1939\)](#), les processus de localisation des entreprises et des individus caractérisés par leur niveau de richesse ne se font pas selon des logiques concentriques, mais selon des logiques sectorielles (figure 1.3). L'espace urbain se découperait alors en secteurs radiant qui se développent le long des principaux axes de transport. Le centre de la ville reste occupé par le CBD, regroupant les activités de petit commerce et de service. Mais les activités économiques comme le commerce de gros et les industries sortent du CBD pour se localiser à proximité des réseaux de transport et des voies ferrées plus particulièrement.

A partir de là, les individus les plus aisées, qui disposent de plus de marge de manœuvre, effectuent leur choix en fonction des implantations industrielles, des caractéristiques du site de la ville et des réseaux de transport. Les classes aisées s'installent sur un côté du CBD, à l'opposé des zones industrielles et de commerce de gros. Puisqu'ils peuvent choisir plus librement leur implantation, les ménages les plus riches cherchent à s'implanter dans les parties les plus favorables de la ville, c'est-à-dire sur des sites présentant des paysages attrayants, sur des sites élevés, le long de cours d'eau non utilisés par l'industrie et surtout à proximité des voies rapides de transport. Dans ce modèle, les localisations des autres individus s'apprécient par rapport aux localisations

initiales des entreprises et des zones résidentielles aisées. Ces dernières sont bordées de zones résidentielles de niveau intermédiaire, tandis que les zones résidentielles des classes les plus pauvres occupent les espaces libres restants.

Comme le modèle de Burgess avant lui, le modèle de Hoyt possède une dimension dynamique. Avec le temps, les localisations des individus évoluent et s'opèrent selon un mouvement d'extension vers la périphérie. Les secteurs riches mais anciens sont progressivement abandonnés et convertis en zones résidentielles intermédiaires et pauvres. Sachant que les secteurs riches sont déjà bordés par des secteurs intermédiaires, le seul mouvement possible est en direction de la périphérie en suivant les aménités environnementales et les réseaux de transport. Le même processus d'abandon des logements anciens et de mouvement vers la périphérie opère pour les zones intermédiaires. Mais les secteurs pauvres s'étendent à la fois selon une progression centripète sur les quartiers dégradés et selon un mouvement centrifuge.



- 1 : Centre des affaires (CBD)
- 2 : Commerce de gros et industries
- 3 : Zone résidentielle des classes pauvres
- 4 : Zone résidentielle des classes moyennes
- 5 : Zone résidentielle des classes aisées

FIGURE 1.3 – Modèle sectoriel de Hoyt, Source : [Harris et Ullman \(1945\)](#)

Construit sur de vastes travaux empiriques menés sur les agglomérations américaines, ce modèle a été critiqué pour son approche plus schématique que modélisatrice, qui rend les villes plus difficilement comparables entre elles ([Racine, 1971](#)). La prise en compte des caractéristiques de site, pourtant indispensable pour comprendre les logiques d'implantation, oblige à intégrer des particularités régionales et donc à sortir du cadre généraliste usuellement associé à la modélisation. Les processus de développement des villes tels qu'ils sont explicités par Hoyt n'intègrent pas les phénomènes de gentrification qui sont observables aussi bien dans des villes américaines, qu'euro-

peennnes depuis les années 1960-70 (Lebreton et Mougél, 2008). Le mouvement vers la périphérie est toujours aussi indéniable aujourd’hui que dans les années 1940, mais c’est sans compter avec le retour vers le centre des classes aisées, qui s’installent dans des quartiers pauvres réhabilités ou des quartiers anciennement industriels transformés par la volonté du pouvoir publique ou par des promoteurs. Cette nouvelle tendance modifie donc l’organisation des secteurs qui avait été identifiée par Hoyt.

Harris et Ullman La première modélisation de la ville comme objet polycentrique est elle aussi réalisée par des géographes de l’école de Chicago, Chauncy Harris et Edward Ullman. En se basant sur les théories de leurs prédécesseurs, Burgess et Hoyt, ils développent un modèle pour décrire les structures urbaines qui admettent plusieurs noyaux polarisateurs (Harris et Ullman, 1945). Un noyau représente alors un espace attractif qui concentre un pouvoir, économique ou politique, dans un espace urbain marqué par la spécialisation. Dans cette vision de l’espace urbain, l’existence du CBD n’est pas remise en cause. Il n’est juste plus nécessairement l’unique espace attractif.

L’émergence de multiples noyaux dépend de l’histoire des villes, et peut se faire de façon simultanée, ou lors de la croissance de la ville à la faveur des processus de migration et de spécialisation. La taille de la ville est un facteur déterminant de l’apparition de plusieurs noyaux spécialisés. Les auteurs résument l’émergence différenciée de ces noyaux en une combinaison de quatre facteurs :

1. Le besoin d’équipements spécifiques : les commerces de détail auront ainsi besoin de l’espace le plus accessible et seront donc localisés dans le CBD ou à proximité des consommateurs. Les industries de production préféreront, elles, les espaces assurant un accès aux dessertes fluviales, maritimes ou ferroviaires.
2. Les externalités et économies d’agglomération : les noyaux émergent des processus de concentration entraînés par ces facteurs économiques, qui encouragent les commerces à se regrouper au même endroit pour attirer plus de clients, ou les entreprises à se regrouper pour les facilités de communication.
3. Les antagonismes de développement : les dégradations environnementales, généralement associées à l’industrie lourde, repoussent les zones résidentielles aisées et les commerces qui leur sont associés. De plus, dans un espace déjà densément construit, l’implantation de nouvelles entreprises ou d’équipements majeurs (quartiers de la gare par exemple) requérant de l’espace ne peut se faire que par dédoublement des noyaux.
4. Le manque de moyen foncier : certains agents n’ont pas les moyens de s’offrir les sites les plus désirables et centraux, que ce soit les commerces de gros ou les résidences populaires.

De ces facteurs découle une organisation spatiale en quartiers-types qui se retrouve d’après les auteurs dans la majorité des grandes villes américaines (figure 1.4). Ce

découpage contient plusieurs centres d'activités économiques : le CBD (1) qui reste la zone la plus centrale vers laquelle convergent tous les moyens de transport ; une zone de commerce de gros (2) à proximité du CBD qui jouit à la fois des infrastructures de transport et de la proximité au marché et à la main-d'œuvre ; une zone concentrant les industries lourdes (6) située à une limite actuelle ou ancienne de la ville ; une zone d'affaires secondaire (7) qui a suivi les migrations des classes aisées et moyennes ; et enfin une banlieue industrielle (9) qui correspond à un espace périphérique déconnecté de l'agglomération centrale mais qui ne pourrait exister sans sa présence. La localisation des zones résidentielles, subdivisées en trois classes de revenus (3,4,5), est conforme aux hypothèses effectuées dans le modèle sectoriel de Hoyt. Ce nouveau découpage intègre, en plus de ces zones résidentielles situées dans l'agglomération centrale, des zones résidentielles dites de banlieue ou satellites (8). Ces espaces en grande périphérie sont liés fonctionnellement à l'agglomération centrale, puisqu'ils sont à l'origine de nombreuses navettes quotidiennes de travail, ou possèdent des activités économiques étroitement liées à la zone centrale.

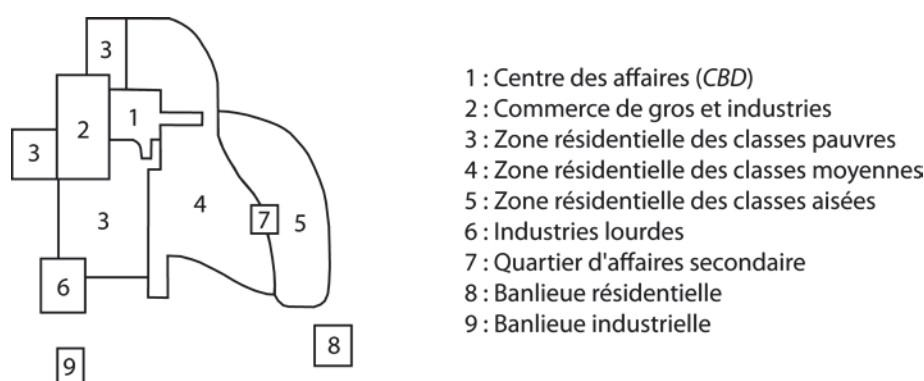


FIGURE 1.4 – Modèle des noyaux multiples de Harris et Ullman, Source : [Harris et Ullman \(1945\)](#)

Le modèle d'Ullman et Harris prend donc une dimension beaucoup plus fonctionnelle en intégrant dans une même structure urbaine des espaces qui ne sont pas morphologiquement liés entre eux. La révolution automobile déjà enclenchée dans les années 1940 aux États-Unis pousse les modélisateurs à intégrer dans l'espace urbain les banlieues des navetteurs qui figuraient déjà dans le modèle de Burgess, mais correspondaient à un espace complémentaire à la ville.

Comme pour le modèle de Hoyt, le modèle de Ullmann et Harris est considéré par certains comme une schématisation de l'espace urbain, plus que comme une modélisation en tant que telle ([Racine, 1971](#)). Bien que leurs travaux soient justifiés par un ensemble de règles, ils se posent comme une description complète mais simplifiée et spatialisée des agencements urbains. Ce modèle aussi schématique qu'il soit, fournit un alphabet de secteurs élémentaires qui vont se composer pour former ce qui apparaît comme la structure moderne de la ville.

1.1.3.2 La différenciation de l'espace dans la géographie structurale

La sectorisation de l'espace urbain est aussi une caractéristique prise en compte dans l'approche de la géographie structurale [Ritchot \(1992\)](#). Dans ce cadre, la ville s'organise et se développe selon des processus de long terme d'appropriation territoriale sous contrôle politique et d'occupation spatiale. Il émerge de ce processus de morphogenèse une structure d'interface géographique et politique qui canalise les forces de valorisation économique et la pression démographique. L'instance politique n'y apparaît pas seulement comme une puissance régulatrice. Elle est garante de la contrainte fondamentale qui régit le rapport entre l'homme et la nature : l'interdit de propriété qui est « la condition de possibilité de l'appropriation, de la transformation et de l'utilisation des ressources du monde extérieur » ([Desmarais et Ritchot, 2000](#), p. 17).

L'appropriation de l'espace urbain est en effet un phénomène de très long terme qui n'a longtemps concerné qu'un petit nombre d'acteurs. L'origine de cette appropriation, et donc l'origine des villes, est théorisée comme l'application de l'interdit de propriété dans un lieu vide, sacré et investi de valeurs anthropologiques, appelés *vacuum*. Le *vacuum* est donc un lieu attractif, dont il est intéressant de contrôler l'accès en occupant des positions proches, puisque l'établissement permanent y est interdit. La ville peut donc se développer à partir de ces positions, que la proximité du *vacuum* valorise, et qu'il faut rentabiliser par une activité économique.

Cette approche contraste avec celles des modèles économiques qui postulent que l'accès aux ressources de la nature est a priori permis et n'est limité que par le pouvoir physique ou financier d'acteurs en compétition. Sans nier la compétition, la géographie structurale affirme que celle-ci est conditionnée par une structure morphologique distribuant des valeurs abstraites dans l'espace, comme par exemple : autorité, fécondité, santé, productivité ou encore connaissance, etc. qui peuvent caractériser certains quartiers d'une ville et orienter leur valorisation économique.

Les dynamiques d'appropriation du territoire sont analysées en tant que trajectoires d'établissement ou de mobilité d'une position à une autre au sein de cet interface. Ce sont les différentes trajectoires de mobilité réalisées au cours de l'histoire qui, caractérisant un ensemble de parcours-types, aboutissent à la sectorisation de l'espace géographique. Puisque l'établissement en position terminale de la trajectoire est régulé par la propriété, il existe un processus de contrôle politique, qu'un acteur peut satisfaire ou pas. La trajectoire d'établissement peut donc être réalisée avec ou sans le contrôle politique de la mobilité. La géographie structurale distingue deux formes de trajectoires : les trajectoires endorégulées et les trajectoires exorégulées. La mobilité exorégulée est celle des individus qui pour avoir accès aux valeurs urbaines doivent accepter d'occuper une position moins valorisée, séparée des lieux de valeur par des barrières sociales ou de distance. A l'opposé, la mobilité endorégulée est celle des individus qui peuvent occuper les positions investies des valeurs visées et ainsi choisir la façon de les rentabiliser.

économiquement. Notons que ces trajectoires peuvent concerner autant des ménages cherchant un logement que des entrepreneurs voulant implanter une activité. Ces deux modalités du contrôle politique de la mobilité, associées à la direction convergente ou divergente des trajectoires, définissent alors un ensemble de valeurs positionnelles (ou parcours-types) :

- le rassemblement (R) : résultat d'une trajectoire « endorégulée focalisante »
- l'évasion (E) : résultat d'une trajectoire « endorégulée diffusante »
- la concentration (C) : résultat d'une trajectoire « exorégulée focalisante »
- la dispersion (D) : résultat d'une trajectoire « exorégulée diffusante »

Ces valeurs positionnelles combinées aux déclencheurs que sont les deux facteurs pression démographique et valorisation économique par la rente justifient la production d'espaces aux aménagements différenciés : « les trajectoires de mobilité engendrent des discontinuités qui catégorisent l'espace géographique en domaines qualitativement distincts » (Ritchot, 1992, p. 195). Ces domaines qualitativement distincts correspondent

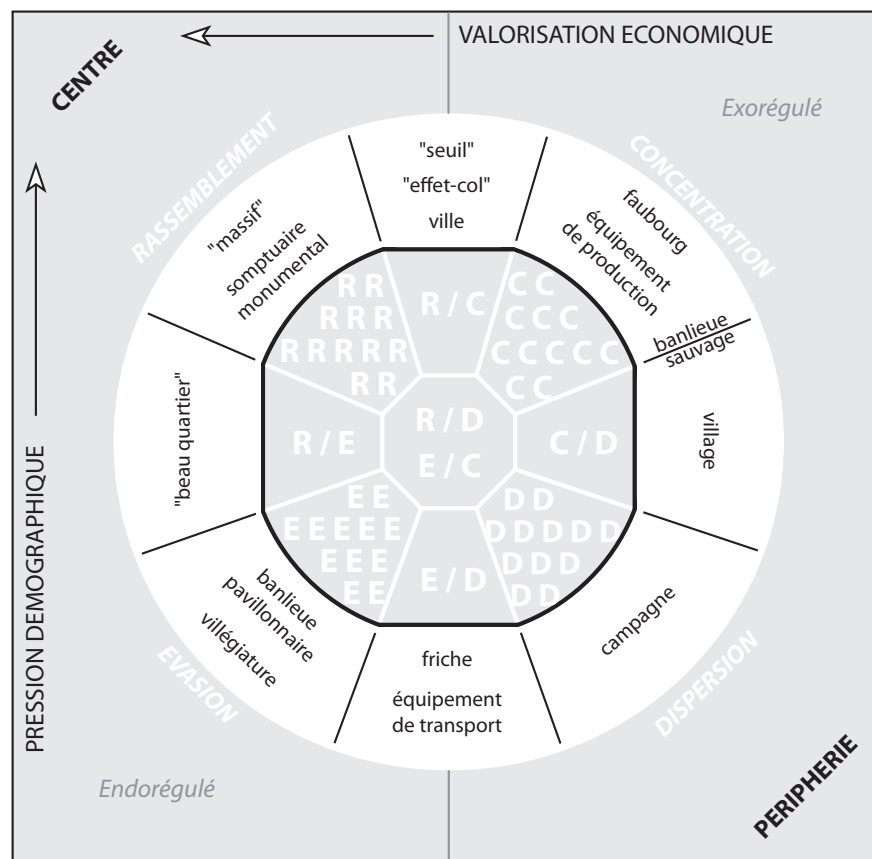


FIGURE 1.5 – Modèle d'occupation spatiale issu de la géographie structurale, Source : Ritchot (1992)

à des quartiers aux niveaux de richesses et aux aménagements différents.

La figure 1.5 présente un résumé de cette théorie l'émergence de la ville par conversion des valeurs positionnelles (R/E/C/D) sous l'action de la pression démographique et de la valorisation économique, en domaines qualitativement distincts. Cette figure montre que les facteurs « pression démographique » et « valorisation économique » conditionnent l'organisation des espaces urbains. Une succession de phases marquées par un type de trajectoire dominant matérialise alors un ensemble de domaines distincts dont la qualification coïncide avec le vocabulaire archétypique des aménageurs, urbanistes et géographes : beaux quartiers, faubourg, banlieue, banlieue pavillonnaire ... Ces quartiers correspondent à une spécialisation des unités de voisinage qui s'opère de façon dynamique selon un enchaînement de phases d'occupation spatiale dont la régulation relève *a priori* du pouvoir politique et *a posteriori* d'une valorisation économique. Différents types d'aménagements se matérialisent selon les phases de nature endorégulée ou exorégulée et de direction focalisante ou diffusante, mais aussi selon la position spatiale de l'unité de voisinage par rapport à la forme abstraite que représente le vacuum et donc au niveau de pression démographique et économique.

1.1.4 La multipolarisation

Le centre urbain, lieu unique qui « articule les pouvoirs, symbolise les lieux, et cristallise les fonctions économiques » (Gaschet et Lacour, 2002, p. 50), devient multiple dans la structure moderne de la ville. La centralité avec son rôle structurant persiste mais la définition du centre change dans ces espaces où fonctions et pouvoirs se délocalisent en des lieux différenciés. Le modèle centre-périphérie évolue et intègre maintenant un ensemble de lieux qu'il faut caractériser par des degrés variables et complémentaires de centralité, et qui forment un réseau.

Nous verrons ici que de nombreuses études récentes ont cherché à définir et à caractériser la composition de ces centres. Les éléments d'analyse présentés ici sont associés à l'utilisation de données spécifiques relatives à l'emploi ou à la mobilité, non disponibles sur les grilles régulières que nous utiliserons. Ils représentent toutefois une caractéristique marquante des structures urbaines que nous tenterons d'approcher à partir d'autres données relatives au cadre bâti qui permettront de révéler les formes inscrites dans la longue durée.

1.1.4.1 Définition des centres

Observer la nature polycentrique des structures urbaines oblige à poser une définition au moins partiellement statistique qui permette d'identifier et de délimiter les centres. Les études fondatrices comme celles de McDonald (1987) ou McDonald et McMillen (1990), basées sur des travaux empiriques concernant des métropoles amé-

ricaines, introduisent le concept de sous-centres d'emplois. En restant dans le cadre polycentrique vu comme un objet de l'économie néo-classique, il apparaît naturel et plus facile au niveau de la disponibilité des données, de chercher les nouveaux centres comme des lieux de concentration d'emplois.

Un sous-centre est alors défini par McDonald comme un site possédant une densité d'emplois significativement plus élevée que dans tous les sites voisins. Dans l'idée que les formes urbaines ont évolué d'une structure monocentrique vers une structure polycentrique, le CBD, élément central et facilement identifiable, conserve un statut à part de centre principal, tandis que les nouveaux ou autres centres d'emplois de moindre taille sont qualifiés de sous-centres ou centres secondaires. Même si de nombreuses études confirment le pouvoir déclinant du CBD dans des formes où les sous-centres prennent de plus en plus d'importance, les deux types restent généralement distincts pour mieux cerner l'aspect polarisateur des nouveaux centres.

La mise en œuvre de la définition de McDonald (1987) présente l'inconvénient d'identifier de petites zones dans la forme urbaine qui ne sont pas spécialement significatives puisque existant par le fait qu'elles sont entourées de zones sans emplois. Les travaux de Giuliano et Small (1991) ont donc repris partiellement cette définition, en ajoutant deux conditions de seuils minima sur la densité d'emplois (10 emplois/hectare) et sur le nombre total d'emplois (10 000 emplois) que doit contenir un sous-centre. L'étude des plus grands centres d'emplois permet de se focaliser uniquement sur les centres assez larges pour pouvoir influencer et polariser l'espace environnant, que ce soit en termes de rente foncière ou de forme urbaine (Giuliano et al., 2012). Bien que parfois critiquée pour la nature arbitraire des seuils (Redfearn, 2007), cette définition simple qui permet d'identifier directement les sous-centres d'emplois d'une agglomération reste l'une des plus reprises (Cladera et al., 2009). Ce qui apparaissait comme arbitraire permet en fait une adaptabilité aux particularités locales qui fait de cette définition une approche très robuste.

Mais au-delà de cette délimitation par seuils, il existe un ensemble foisonnant de méthodes pour identifier des sous-centres d'emplois. A ce titre, l'article de Cladera et al. (2009) fournit un état de l'art assez détaillé des approches communément employées. Il existe ainsi un vaste corpus de méthodes paramétriques et non paramétriques pour identifier des résidus ou des pics sur la densité d'emplois. Mais comme le remarquent les auteurs, le problème de ces approches reste la difficulté à identifier de vrais centres parmi ces pics et résidus, qui ne soient pas juste le résultat d'un processus aléatoire dans le développement urbain. Pour vérifier que ces pics correspondent bien à des éléments structurant l'objet urbain, certaines études se basent sur l'analyse des flux de mobilité domicile-travail, ou de mobilité motivée par des fins économiques plutôt que sur la densité d'emplois. Les sous-centres délimités peuvent alors répondre aux critères qui déterminent leur existence, à savoir des concentrations d'entreprises assez importantes pour influencer l'organisation spatiale environnante.

1.1.4.2 Caractéristique des centres

Comme le montraient déjà les modèles de la première génération, comme celui de Hoyt ou de Harris et Ullman, la décentralisation des activités initialement dans le CBD ne se fait pas de manière homogène. Les premières activités à avoir impulsé de nouveaux *clusters* sont les activités fortement consommatrices d'espace, comme l'industrie lourde ou le commerce de gros. D'autres phases de décentralisation ont suivi (Gaschet, 2002), concernant cette fois le commerce de détail avec l'apparition de centres commerciaux par exemple, puis les activités tertiaires supérieures telles que les sièges sociaux, activités financières ou de communication jusque-là restées attachées aux CBD. L'apparition de technopoles constitue l'exemple marquant de la fuite du CBD des activités de haut niveau. Ces phases de décentralisation correspondent surtout à des phases de spécialisations spatiales de certaines activités qui produisent les centres d'une structure polycentrique, et qui peuvent se révéler très hétérogènes, que ce soit en termes de forme ou de contenu.

Même si les emplois quittent le CBD pour former de nouveaux centres ou pour se disperser, l'influence et le poids du CBD restent dominants, même dans les agglomérations américaines (Anas et al., 1998). Le poids du centre historique reste écrasant dans la majorité des aires urbaines françaises où il représente de loin le pôle contenant le plus d'emplois et d'activités (Mignot et al., 2007).

Aux Etats-Unis, cette décentralisation des activités économiques a été accompagnée par un grand phénomène de dispersion des activités. A Los Angeles, métropole caractérisée par un fort étalement urbain, seulement un tiers des emplois se localisent dans des centres (Giuliano et Small, 1991), qui ne représentent que 3% du territoire de la métropole. En comparaison, Cervero et Wu (1997) notent que les centres d'emplois de San Francisco accueillent 47% des emplois de la métropole. La situation en Europe, et plus particulièrement en France apparaît tout à fait différente puisque les emplois restent majoritairement concentrés. Mignot et al. (2007) montrent que l'emploi est concentré à environ 90% dans des pôles pour les aires urbaines de Lyon, Marseille et Lille. Ces différences sont bien sûr imputables aux trajectoires urbaines très différentes sur les deux continents, mais la différence de référentiel de données tient aussi un rôle important. L'étude sur Los Angeles est réalisée à partir de '*transportation analysis zones*' équivalentes aux secteurs de tirage des enquêtes ménages-déplacements, celle portant sur la Baie de San Francisco est réalisée à partir de secteurs de recensement, tandis que l'étude des aires urbaines françaises est réalisée à partir du découpage communal.

Bien que les études citées se basent sur des maillages hétérogènes et des données différentes, il ressort dans le cas américain comme dans le cas français que les centres d'emplois correspondent en fait à des zones relativement mixtes en population et emplois (Giuliano et Small, 1991; Mignot et al., 2004). Seuls des centres particuliers comme les zones aéroportuaires possèdent une très forte spécialisation fonctionnelle puisque

pratiquement aucune population n'y réside. La spécialisation en termes de secteurs d'activité peut par exemple se traduire par l'existence de *clusters* de services (médicaux ou universitaires notamment). Mais il existe aussi des centres plus mixtes où activités de production et de service se côtoient avec une dominance des services. Dans l'exemple de Los Angeles, la taille de l'aire métropolitaine fait que certains des sous-centres identifiés correspondent en fait aux CBD plus anciens de villes qui étaient précédemment indépendantes morphologiquement et/ou fonctionnellement. Malgré cette particularité propre aux grandes aires urbaines, [Gaschet et Lacour \(2002\)](#) notent qu'une des caractéristiques fortes du polycentrisme est la spécialisation des sous-centres qui généralement ne contiennent qu'un ou deux types d'activités économiques.

Morphologiquement parlant, les sous-centres restent clairement distincts du centre principal lorsqu'ils se développent à la périphérie de l'agglomération. Mais comme les sous-centres d'emplois ont tendance à se développer en privilégiant la proximité aux axes de transport, il apparaît dans des villes comme Los Angeles ([Giuliano et Small, 1991](#)) ou Houston ([Anas et al., 1998](#)) que les sous-centres forment des corridors d'emplois connectés au CBD qui contiennent une part importante de l'emploi métropolitain. Cette forme de diffusion progressive des activités hors du CBD révèle l'importance de l'échelle et du référentiel de données utilisé pour les sous-centres, qui bien que proches peuvent se différencier par leur composition.

Les multiples sous-centres des structures polycentriques poussent à faire évoluer la notion de centralité qui est maintenant assurée par un système de pôles gravitant autour du centre-ville historique. Ce système de centres plus ou moins spécialisés a conduit des auteurs tels que [Heikkila et al. \(1989\)](#) à envisager l'existence à l'échelle régionale de plusieurs formes de polycentrismes pour modéliser la valeur des terrains dans l'espace urbain. Ils distinguent alors trois types de fonctionnement polycentrique, caractérisées par des centres qui peuvent être complémentaires ou substituables :

1. *Polycentrisme aux centres complémentaires* : l'accessibilité en chaque point de l'espace est valorisée par rapport à tous les centres.
2. *Polycentrisme aux centres substituables* : l'accessibilité n'est valorisée que par rapport au centre le plus influent, chaque centre polarise son espace local indépendamment.
3. *Polycentrisme aux centres partiellement complémentaires et substituables* : l'accessibilité est valorisée par rapport à tous les centres mais avec une influence moins forte des centres les plus éloignés.

Le modèle polycentrique correspond en fait à un ensemble varié de configurations, dépendant autant de la taille de la ville et donc du nombre de centres, que du poids plus ou moins important du centre principal, des caractéristiques des sous-centres et de leur fonctionnement en tant que système.

1.1.4.3 Évolution des sous-centres

Un des constats marquants de l'évolution des formes urbaines qualifiées de polycentriques est que les sous-centres distincts du noyau traditionnel grandissent plus vite que ce dernier. Les résultats de différentes études sont parfois contradictoires sur les tendances de croissance, mais il semble indéniable que le centre historique perd progressivement son influence au profit des sous-centres existants ou émergents (Mignot et al., 2004; Giuliano et Redfearn, 2005).

Inscrites dans le débat lancé avec les travaux de Gordon et Richardson (1996) portant sur la structure de Los Angeles, de nombreuses études sur l'évolution des structures polycentriques relativisent le poids du polycentrisme face aux phénomènes d'étalement et de dispersion. L'étalement de la population urbaine observable sur tous les continents a été suivi, plus particulièrement aux Etats-Unis, par un étalement des activités économiques qui fait que les centres et sous-centres d'emplois ne contiennent plus la majorité des emplois des régions métropolitaines. Gordon et Richardson (1996) montrent ainsi que l'emploi décline aussi bien dans le centre principal que dans les sous-centres de Los Angeles. Giuliano et Redfearn (2005) confirment partiellement cette tendance décroissante en la contrastant. La part d'emplois du *Downtown* et de certains sous-centres périphériques diminue. Mais certains sous-centres de l'agglomération principale de Los Angeles sont stables, et d'autres sous-centres des banlieues les plus anciennes connaissent une nette tendance à l'augmentation de la concentration d'emplois. Ainsi même si la dispersion est clairement observable à la périphérie où les *edge cities* (Garreau, 1991) peuvent laisser la place aux *edgeless cities* (Lang et LeFurgy, 2003), cette tendance n'est pas globale. Et comme l'explique Anas et al. (1998), si les travaux de Gordon et Richardson montrent bien une augmentation de la dispersion à un instant donné, ils ne montrent pas si cette tendance continue à produire des forces agglomérantes qui à terme produiront de nouveaux sous-centres.

Ces trajectoires hétérogènes des centres d'emplois ont particulièrement été étudiées par Giuliano et Redfearn qui ont cherché à identifier les facteurs qui orientaient l'évolution des sous-centres d'emplois de Los Angeles. Dans une première étude, Redfearn et al. (2008) constatent que les sous-centres les plus petits et les moins denses sont les plus à même de grandir plus rapidement puisqu'ils disposent de plus de terrains disponibles. En comparaison, l'étude française de Mignot et al. (2004) fait le constat inverse puisque ce sont les sous-centres des banlieues déjà fortement urbanisées qui connaissent la croissance la plus importante. Ceci démontre une fois de plus les différences structurelles des agglomérations américaines et européennes.

De plus, la croissance semble liée aux activités présentes dans les sous-centres, les sous-centres associés aux secteurs d'activités les plus porteurs étant caractérisés par une croissance plus importante. L'accessibilité ayant toujours été pointée comme un des facteurs les plus importants, Giuliano et al. (2012) notent dans leur synthèse que

si le réseau routier a un effet positif sur l'émergence des sous-centres, il n'a pas de rôle significatif dans leur croissance. Parmi les différentes variables d'accessibilité testées, celle qui semble avoir le plus d'impact sur leur croissance est l'accès à la force de travail. La recherche d'accessibilité aux réseaux de transport ou à la force de travail, le rôle des politiques d'aménagement et des promoteurs privés, les volontés contradictoires entre recherche de parcelles bon marché et recherche de positions de prestige sont autant de facteurs qui vont jouer sur l'évolution des localisations des entreprises et des centres d'emplois.

1.1.5 Les systèmes de villes

La variété observée des structures polycentriques révèle des centres et sous-centres d'emplois de spécialité, de taille et d'influence différentes. Les structures polycentriques des plus grandes métropoles telles que Los Angeles ou Paris exercent leur pouvoir de centralité sur un système de centres qui contient aussi bien de nouveaux sous-centres spécialisés que d'anciens centres-villes de structures urbaines anciennement indépendantes. Ces exemples, qui morphologiquement parlant tiennent plutôt de la région urbaine agglomérée et polarisée, rappellent que la compréhension des structures urbaines ne peut se faire qu'en articulant les échelles d'études. Les objets urbains métropolitains sont façonnés par des interactions à différentes échelles qui influencent aussi bien la forme que la fonction. Les travaux sur le polycentrisme exposés dans la partie précédente, attachés aux approches micro-économiques correspondaient à des études à l'échelle de l'agglomération, échelle considérablement variable compte tenu des étendues différentes des métropoles citées. La complexité des relations et des formes de polycentrisme déjà observées pousse à forcer le changement d'échelle et à mettre en regard ce modèle général avec les théories, plus propres à la géographie, des systèmes régionaux de villes.

1.1.5.1 La théorie des lieux centraux de Christaller

A l'échelle du système de villes, la théorie des lieux centraux développée par **Christaller (1933)** continue de faire référence parmi les géographes. Les travaux de ce géographe allemand voulaient expliquer la taille, le nombre et la répartition des villes en Allemagne du Sud dans un cadre de travail situé entre la géographie et l'économie spatiale. Dans cette théorie, la ville est vue comme un lieu central créateur et consommateur de biens et services, mais aussi un centre de redistribution qui fournit biens et services à tout un territoire environnant, définissant ainsi une aire d'influence.

Cette théorie axée sur la centralité des lieux correspond à une vision de la ville comme pôle élémentaire d'une vaste structure de réseau de villes. Elle correspond au principe repris par des géographes comme **Berry (1964)**, **Pumain (2006a)** ou **Bretagnolle et al. (2009)**, qui veut qu'on ne peut comprendre un objet urbain sans le réintégrer dans

le système de peuplement plus global dont il fait partie : « A ce titre, une ville n'apparaît jamais isolément, elle fait toujours partie d'un réseau de villes, elle est un nœud dans un système de peuplement, qui évolue de manière cohérente, en phase avec les autres éléments de ce système » (Pumain, 2006a, p. 36).

La théorie des lieux centraux repose ainsi sur les hypothèses suivantes :

- L'espace du système de villes est non fermé, isotrope et homogène.
- L'espace est régi par les lois économiques de l'offre et de la demande.
- Les coûts de transport influent sur les choix des consommateurs : ils cherchent à minimiser le coût et donc la distance en achetant au centre le plus proche.

L'espace s'organise alors en un système de lieux centraux caractérisés par leurs aires de marchés qui sont, elles, déterminées par des effets de seuil et de portée. Les seuils concernent la taille minimale que doit avoir un marché, en termes de population dans le modèle de Christaller, pour que certains biens et services apparaissent et soient rentables. La portée concerne, elle, la distance maximale qu'un consommateur accepte de parcourir pour se procurer un bien ou un service donnés. De là découle une notion de hiérarchie urbaine basée sur les postulats qui énoncent que les biens et services de même portée se regroupent dans les mêmes centres, et que les centres qui fournissent les biens et services d'une certaine portée englobent nécessairement ceux de portée inférieure (Pumain et al., 2006). Ainsi, plus une ville a une population importante, plus elle offre de biens et services variés et rares, et plus son aire d'influence sera importante. Le réseau de villes est donc hiérarchisé en fonction de la population des villes, et par suite en fonction du niveau de rareté des biens et services qu'elles procurent à leur aire d'influence.

En partant de ces hypothèses, Christaller construit une organisation spatiale en aires de marchés triangulaires qui une fois combinées forment un treillis hexagonal, où les lieux centraux sont régulièrement espacés. Cette organisation hexagonale représente le pavage optimal en termes de minimisation des coûts de déplacement et de gestion des frontières territoriales (Haggett, 1973). Les lieux centraux sont alors organisés selon une hiérarchie spatiale qui entraîne que les centres les plus importants en population, c'est-à-dire de plus haut niveau, sont plus espacés que les centres les plus petits, et que tous les centres de même niveau sont équidistants. Christaller observe alors que plusieurs configurations spatiales sont possibles (figure 1.6) et régies par trois grands principes : le principe de marché, le principe de transport et le principe d'administration. Dans un espace fonctionnant selon le principe de marché, la configuration optimale minimise le nombre de points de production. Une aire de marché hexagonale correspond ainsi à un centre principal au centre de l'hexagone et six centres secondaires situés sur les sommets de l'hexagone. Dans un espace fonctionnant selon le principe de transport, la recherche de la minimisation des coûts de transport se traduit par une organisation

où les centres secondaires ne sont plus sur les sommets de l'hexagone mais au milieu des arêtes. Enfin dans un espace fonctionnant selon le principe d'administration, les centres secondaires se situent à égale distance du centre principal qui les administre, dans l'hexagone. Principalement en réponse à ce dernier principe, Christaller établit pour le réseau de villes d'Allemagne du Sud une typologie des villes selon leur niveau administratif (capitale, chef-lieu de province, de district, etc.).

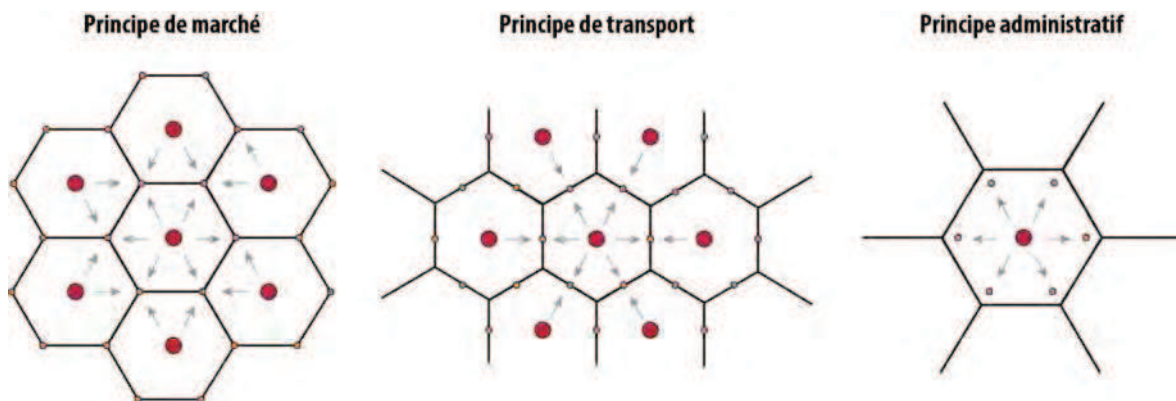


FIGURE 1.6 – Modèles d'organisation des lieux centraux de Christaller

Les travaux de l'économiste allemand **Lösch (1954)** constituent une extension micro-économique des travaux de Christaller. L'approche est beaucoup plus axée sur les questions d'équilibre économique des systèmes de lieux centraux. Les approches en partie différentes de Christaller et Lösch proposent toutes les deux une conception similaire des villes vues à travers le prisme de la centralité, des hypothèses communes et une modélisation des systèmes de villes sous forme de treillis hexagonaux. A l'opposé de Christaller qui partait des villes de plus haut niveau pour construire son système, Lösch modélise l'organisation de l'espace en partant du plus bas niveau constitué d'entreprises autonomes régulièrement espacées. De là, en supposant que les entreprises sont libres d'entrer dans le système dans le but de maximiser leur profit, Lösch montre mathématiquement que les configurations d'équilibre économique correspondent à des arrangements hexagonaux des lieux centraux. Son modèle est plus général puisqu'il ne travaille pas exclusivement sur des hiérarchies inclusives de centres. Mais parmi les configurations correspondant à des situations d'équilibre économique, Lösch retrouve les trois organisations spatiales identifiées par Christaller.

Par rapport aux travaux de Christaller, les travaux de Lösch présentent l'intérêt de pouvoir tenir compte de la spécialisation économique des lieux centraux. En effet, certaines villes marquées par une spécialisation tournée par exemple vers l'exploitation industrielle de ressources naturelles, ou encore vers le tourisme n'entrent pas dans l'hypothèse de Christaller de localisation à proximité de la clientèle. La spécialisation peut soutenir une croissance de population urbaine importante, mais elle n'est pas nécessairement accompagnée d'une diversification des services et de l'apparition de

services de haut niveau. Les villes minières du Nord de la France constituent un exemple de cette spécialisation qui entraîne une concentration importante de la population mais sans l'apparition de tous les services attendus à leur niveau dans la hiérarchie urbaine française (Pumain, 2006b).

La théorie des lieux centraux, bien que plébiscitée pour comprendre l'organisation spatiale des activités a parfois été critiquée notamment sur l'hypothèse selon laquelle les consommateurs cherchent à acheter au plus proche. Cette hypothèse semble invalide dans les formes urbaines modernes, où les consommateurs privilégient les déplacements à buts multiples généralement dans un centre de haut niveau qui n'est pas nécessairement le plus proche (Pumain et al., 2006). Cette théorie fournit une vision statique des systèmes urbains, mais à l'heure où les évolutions technologiques tendent à abolir la distance euclidienne, il n'est pas garanti que ces treillis hexagonaux, déjà difficiles à associer à des configurations urbaines réelles, restent des configurations optimales.

La théorie des lieux centraux trouve son intérêt dans la mise en avant de la centralité et plus particulièrement des relations systémiques et hiérarchiques entre différents lieux de centralité. Cette focalisation sur la centralité pousse à la mettre en relation avec les formes qualifiées de polycentriques, au sens exposé dans la section précédente. Ces organisations reposent sur un fonctionnement commun en réseaux de pôles complémentaires et hiérarchisés, qui ont pour vocation de desservir toute une population urbaine, mais aussi périurbaine et rurale, en biens et services administratifs et économiques. Le polycentrisme de la section précédente est avant tout un modèle fonctionnel qui caractérise la centralité comme une forme de pouvoir économique qui se traduit par une concentration dans l'espace d'activités plus ou moins spécialisées. La place centrale qu'est la ville est, elle, plus qu'un centre d'activités économiques : c'est un centre de commandement qui a la capacité d'influencer le territoire environnant aussi bien sur le plan économique et politique que social.

1.1.5.2 La hiérarchie aplatie : les villes en réseau

La logique d'organisation derrière la théorie des lieux centraux de Christaller repose sur un modèle gravitaire basé sur la substituabilité des centres de même niveau. Cette logique ne tient pas compte de la spécialisation des villes et donc de leur potentielle complémentarité qui peut amener à intégrer d'autres facteurs explicatifs de l'organisation des systèmes urbains. De là est née une nouvelle théorie de l'organisation polycentrique des espaces urbains qualifiée de *network cities*, que nous traduirons par villes en réseaux.

Cette forme d'organisation urbaine étudiée à l'échelle régionale sous le nom de villes en réseau (Batten, 1995) ou de région urbaine polycentrique (Parr, 2004) correspond à une structure morphologique bien particulière. Les auteurs distinguent dans cette catégorie d'organisation spatiale les régions composées de centres urbains voisins de

taille à peu près égale, associés à des villes distinctes morphologiquement. Pour Parr (2004), les régions urbaines polycentriques se distinguent en plus par le fait que les centres urbains doivent être séparés par des terrains ouverts (agricoles ou vacants). Un minimum de séparabilité doit être observé afin de ne pas intégrer dans la même catégorie d'organisation les formes urbaines coalescentes comme la conurbation Lille-Roubaix-Tourcoing ou les aires métropolitaines multicentrées et étalées comme Los Angeles.

La spécificité des villes en réseau se trouve donc dans l'absence de hiérarchie entre les centres urbains qui interagissent en grande partie selon des principes de complémentarité. Les interactions entre les centres sont multidirectionnelles et correspondent à une superposition des marchés d'emplois ou des aires d'influence commerciales Parr (2004). Ainsi le modèle de villes en réseau correspond à une vision de l'espace géographique où ce n'est pas tant la centralité, mais plutôt la nodalité, la position d'une ville dans un réseau qui importe le plus (Meijers, 2007).

Batten (1995) justifie l'existence de ces structures par un renversement de la tendance qui associait innovations et activités basées sur la connaissance aux plus grandes villes des systèmes urbains. Aujourd'hui la taille des villes ne semble plus être un facteur aussi important, et les activités liées à l'innovation se développent dans des villes de plus petite taille. L'internationalisation de l'économie se traduit par une nécessité, pour les firmes multinationales notamment, d'être localisées sur plusieurs marchés et donc de multiplier leurs implantations. Capello (2000) note ainsi que les logiques de compétition entre les villes de taille moyenne s'en trouvent modifiées : pour s'affirmer dans la compétition à longue distance, ces dernières tirent des avantages à coopérer et à fonctionner en réseau synergique. L'objectif n'est alors plus de minimiser les coûts de transports ou de contrôler des aires de marché qui sont maintenant entremêlées, mais d'exploiter les économies d'échelle propres au fonctionnement en réseau.

Le concept de « taille empruntée » (Alonso, 1973), affirme qu'une ville, quelle que soit sa taille, bénéficie d'économies d'agglomération si elle est située à proximité d'autres centres urbains. Appliqué aux villes en réseau, ce concept permet de comprendre que la mise en commun de la population et des moyens de plusieurs villes à proximité leur permet d'atteindre une masse critique plus importante. La mise en commun des marchés aboutit à la création d'un marché unique plus important. Elle permettrait éventuellement d'atteindre les mêmes avantages en termes de production et le même niveau de services qu'une configuration métropolitaine de grande taille.

L'intérêt porté aux configurations de type « villes en réseau » réside ainsi surtout dans le fait qu'elles sont supposées posséder de nombreux avantages par rapport aux configurations monocentriques en termes d'aménagement et de durabilité. L'hypothèse faite est que ces configurations fonctionnant en synergie peuvent bénéficier de toutes les économies d'agglomération et externalités des configurations monocentriques, sans

pour autant subir les déséconomies qui leur sont propres. Ces structures urbaines, possédant une plus grande créativité et diversité (Batten, 1995), seraient moins sujettes à la congestion, au manque d'espace et à la flambée des prix du foncier ou encore à la pollution (Meijers, 2008). Et comme le souligne Batten (1995), les progrès dans les réseaux de transport, même en comptant la contrainte du budget temps de transport des résidents (Zahavi, 1974), offrent plus de possibilités de développement limitant la congestion dans les structures polycentriques en réseau.

Il ressort des études que ces réseaux n'existeraient pas sans la volonté des aménageurs. Ces réseaux correspondent à un idéal urbain, combinant cadre de vie agréable et compétitivité économique. Comme le souligne Batten (1995) à propos du cas de la Randstad hollandaise, s'il n'y avait pas de projet urbain porté par des pouvoirs publics un minimum coordonnés à l'échelle régionale, alors le réseau se transformerait en structure marquée par la coalescence. C'est la volonté de préserver le *Green Hart* qui permet à la Randstad, modèle du genre des villes en réseau, de le rester. Si les villes tirent des avantages économiques à s'associer pour obtenir un rayonnement plus important, elles n'en restent pas moins en compétition entre elles pour se maintenir, et maintenir leur autonomie. La mise en réseau de villes de taille égale ne peut donc se faire sans l'implication des acteurs de l'aménagement de chaque ville (Capello, 2000). La coopération n'est pas sans risques puisque les effets se font sur le long terme, et qu'il est difficile d'assurer l'équité des retours positifs pour tous les centres (Parr, 2004).

Peu d'études ont cherché à montrer que les villes en réseau bénéficiaient réellement d'une situation similaire ou meilleure que celle d'autres formes d'organisation. Ces hypothèses ont été élaborées majoritairement à partir de l'étude des réseaux les plus compétitifs et importants en termes de taille, comme la Randstad hollandaise. Dans ce cadre, Meijers (2008) a réalisé une étude sur l'ensemble des systèmes urbains régionaux hollandais afin de déterminer si les configurations polycentriques étaient réellement avantageuses en termes d'aménités culturelles et de loisirs. Il observe que les systèmes les plus polycentriques possèdent invariablement moins d'aménités que les systèmes marqués par la dominance d'un centre. Il rejoint ainsi les conclusions de Parr (2004) en affirmant que les externalités semblent moins fortes dans les systèmes de villes en réseau qui ne possèdent pas toujours la nature métropolitaine des agglomérations monocentriques associées à une densité et une proximité plus fortes. Meijers (2008) justifie cette absence de gain dans les structures en réseau par une compétition toujours présente entre les villes qui les pousse à dupliquer les aménités de niveau moindre en chaque centre. Pour ces services qui ne relèvent pas entièrement des lois de marché, l'histoire indépendante et l'inertie de chaque ville ne favorise pas spécialement l'émergence de nouvelles infrastructures habituellement soutenue par les pouvoirs publics.

Dans la pratique, l'étude des structures urbaines à l'échelle régionale montre donc une opposition qui n'est que partielle entre des systèmes urbains hiérarchisés, tels que

décrits par Christaller, et les systèmes urbains « horizontaux » fonctionnant en réseau. Les systèmes urbains en réseau observés correspondent en fait à des intermédiaires entre les deux types de configurations théoriques (Capello, 2000; Meijers, 2007). Certaines activités basées sur la connaissance ou les services, hospitaliers par exemple, peuvent maintenant fonctionner dans un réseau complémentaire non hiérarchisé. Mais d'autres services sont toujours régis par une distribution spatiale basée sur la hiérarchie urbaine.

1.1.5.3 Fractalopolis, le modèle de la ville fractale

A côté de ce modèle des hiérarchies aplaties, Frankhauser et al. (2007) proposent une actualisation différente de la théorie de Christaller correspondant au modèle de la ville fractale. Ce modèle développé pour rendre compte du contexte actuel marqué par l'étalement des territoires artificialisés en périphérie des villes, repose sur une vision différente des systèmes urbains. Ici la hiérarchisation des lieux centraux est toujours une caractéristique clé, mais l'organisation spatiale se matérialise différemment. Elle se matérialise selon une logique d'organisation fractale multi-échelle qui fait revoir la spatialisation des centres et de leurs aires d'influence.

Le modèle Fractalopolis (Frankhauser, 2012) que nous décrivons ici est un modèle inscrit dans la lignée des travaux de modélisation de la ville fractale qui ont commencé dans les années 1980 avec les travaux de Mandelbrot (1982), Batty et Longley (1986) ou Frankhauser (1994). On peut aussi renvoyer vers l'article de Tannier (2009) qui dresse un état de l'art de quelques modèles de ville fractale, utilisant différentes formes (tétragone, tapis de Sierpinski) pour expliquer des processus allant de la croissance urbaine vue en termes de consommation d'espaces, à la hiérarchisation des centres de services, en passant par la modélisation de gradients de rente foncière.

La caractéristique commune à ces travaux est l'utilisation de la géométrie fractale, reposant sur l'utilisation de lois de puissance, pour modéliser la relation multi-échelle d'auto-similarité qui existe entre des formes urbaines locales et globales. Dans les espaces urbains observés, cette relation d'auto-similarité se dessine non pas comme une répétition stricte de formes, mais plutôt comme une répétition de caractéristiques numériques (des dimensions fractales). Cette répétition numérique traduit une logique d'organisation plus relâchée que celle des fractales théoriques, une logique qui repose sur la conservation de la disposition des éléments les uns par rapport aux autres mais selon des processus plus aléatoires ou plus complexes qui permettent de générer des formes irrégulières. En fait, le caractère fractal des espaces urbains ne concerne pas tant la répétition de formes bâties que la répétition des espaces vides complémentaires qui restent présents et s'affinent à travers les échelles.

Le modèle Fractalopolis diffère des modèles exposés jusqu'ici puisqu'il est présenté par ses auteurs comme un modèle descriptif, construit à partir d'une analyse des configurations périurbaines et des tissus urbains existants. Mais il est aussi présenté comme

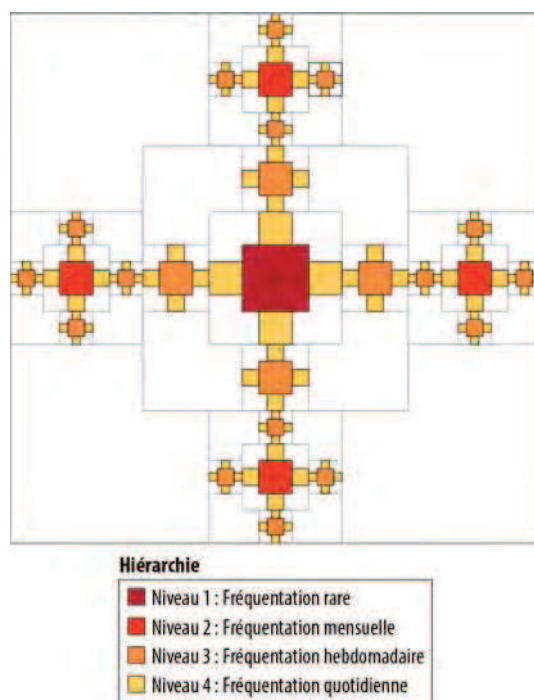


FIGURE 1.7 – Modèle d’organisation Fractalopolis (selon un tapis de Sierpinski multi-fractal)

un modèle normatif, développé pour faire des analyses prospectives et proposer des scénarios d’aménagement dans le but de répondre à des enjeux de durabilités. L’idée est alors de développer un modèle d’organisation répondant aux quatre principes suivants :

1. Réduire les distances parcourues pour accéder à des équipements de haut niveau,
2. Respecter la diversité de la demande sociale en termes de préférences résidentielles et notamment prendre en compte la demande pour un mode de vie périurbain,
3. Éviter l’urbanisation discontinue qui augmente les distances parcourues pour accéder aux centres,
4. Éviter la fragmentation des espaces naturels et agricoles.

Dans cette théorie revisitée, il est supposé préférable de concentrer les centres, et lieux d’habitation associés, à proximité des principaux axes de transport (qui peuvent être interprétés comme des axes de transport public). Cette hypothèse est formulée en lien avec les principes d’aménagement appliqués dans certaines métropoles d’Europe du Nord, et renvoie surtout au principe du TOD (*Transit-Oriented Development*).

Pour modéliser une structure urbaine qui réponde à tous ces principes, il est alors proposé d’utiliser une représentation correspondant à une forme multifractale générée selon le modèle du tapis de Sierpinski, donnée en figure 1.7. Cette structure multi-échelle est un support de modélisation auquel on peut associer quatre types d’informations :

- l’ensemble des carrés colorés qui peuvent être vus comme des aires d’influence, ou

comme les surfaces ouvertes à de futurs développements urbains dans un modèle prospectif ;

- l'ensemble complémentaire des carrés blancs accueillant les espaces naturels et agricoles, les espaces ruraux d'habitat dispersé n'étant pas ouverts aux futurs développements urbains ;
- la hiérarchie des lieux centraux, correspondant aux centroïdes des carrés, distingués en quatre niveaux de services
- la répartition de la population dans les aires d'influence qui se fait, elle aussi, selon un modèle de répartition multifractale.

Cette forme d'organisation possède la propriété d'être une forme optimale en termes d'accessibilité à des aménités urbaines et rurales. Les espaces bâtis sont concentrés autour de nœuds bien desservis par les réseaux de transport ce qui assure une bonne accessibilité au centre. Mais de grands espaces non bâtis pénètrent les tissus urbains, ce qui génère une bordure sinueuse assurant le maintien de nombreux terrains à proximité des aménités rurales. Le développement articulé en branches suivant les ramifications du réseau de transport assure une limitation de la fragmentation des tissus urbains. Cette forme urbaine est alors avantageuse en termes de proposition d'aménagement, puisqu'elle pourrait répondre à plusieurs objectifs de durabilité, en plus de proposer une alternative au scénario largement diffusé de la densification.

Par rapport à la théorie de Christaller, la logique de hiérarchisation des centres est similaire. La logique de hiérarchisation des bassins d'influence n'est toutefois plus la même, puisque la procédure de génération itérative de la forme multifractale peut aboutir à la création d'aires d'influence de taille similaire pour des centres de niveaux différents. Cette caractéristique de la logique multifractale est en fait intéressante pour modéliser le fait que des villes associées à des centres de haut niveau sont souvent plus grandes que celles associées à des centres de bas niveau. Une hiérarchisation complète peut toutefois être réintroduite en utilisant un système de codage binaire pour distinguer des aires d'influence de même surface mais de niveaux hiérarchiques différents (Frankhauser, 2012).

Le même processus itératif de forme multifractale étant appliqué pour générer un modèle de répartition de la population dans les aires d'influence, la même dégénérescence apparaît. La procédure itérative est donc modifiée pour répartir la masse de population (en utilisant une connaissance empirique) de façon à ce que la population soit plus importante en remontant dans la hiérarchie, même si les aires d'influence sont de même taille : les centres les plus importants sont associés à de plus hautes densités de population.

Si ce modèle d'organisation apparaît aussi théorique que celui de Christaller, il présente l'avantage d'être facilement transférable à des organisations urbaines étalées plus

réalistes (sans pertes des propriétés du modèle théorique). Une part d'irrégularité peut être introduite dans la position des carrés générés à chaque itération, sous condition que les carrés soient contenus dans le carré initial, et qu'ils ne se chevauchent pas. Il est alors possible de se rapprocher de configurations plus particulières et de tester leur adéquation au modèle.

1.1.6 Discussion : la centralité et les périmètres urbains

La variété des modèles de configuration polycentrique interroge la caractérisation des structures urbaines et des formes de centralité. Les relations entre les modèles multipolaires ou polycentriques et les systèmes de villes apparaissent complexes puisqu'ils se basent sur les mêmes notions plus ou moins bien définies de centre et centralité, mais sur une échelle d'étude et des périmètres urbains différents.

La centralité La notion de centralité derrière les modèles polycentriques de la section 1.1.4 correspond principalement à une centralité économique qui est associée aux phénomènes d'agglomération d'entreprises. Dans la pratique, la nature polycentrique des villes est donc étudiée à travers l'angle de l'emploi ou sous forme de flux de mobilité domicile-travail. C'est la multiplication et la spécialisation des centres économiques à l'échelle d'une agglomération qui sont observées afin d'évaluer l'influence des différents centres d'activité dans cette agglomération. Dans ce cas, le polycentrisme structure non seulement l'organisation morphologique mais aussi l'organisation fonctionnelle ou relationnelle. Mais sans le centre principal, les espaces polarisés par les sous-centres n'existeraient pas en tant que villes indépendantes.

La centralité des modèles de systèmes urbains de la section 1.1.5 correspond aussi à une centralité économique, mais qui est complétée par une centralité institutionnelle, culturelle et sociale. Ici les villes sont regardées à plus grande échelle et seules les relations entre villes, comme lieux de centralité, comptent. Cette différence de qualification entraîne un polycentrisme qui n'est pas tout à fait le même. A cette échelle le polycentrisme est relationnel, et peut se traduire par la mise en place d'un système hiérarchisé ou au contraire plutôt équilibré. Même si le polycentrisme est relationnel et vu à l'échelle régionale, il impacte aussi la structure locale morphologique et fonctionnelle, ce qui se retrouve dans le modèle Fractalopolis qui constitue un intermédiaire entre les deux catégories de modèles.

La variété et la complexité de ces modèles tient donc en partie à l'imbrication de trois dimensions qui permettent de caractériser l'espace et donc les centres :

- la *dimension morphologique* qui permet de délimiter des centres urbains et des espaces urbains en fonction d'une caractéristique de proximité qui est déterminante pour des questions de développement urbain, de consommation d'espace

et de paysage urbain ;

- la *dimension fonctionnelle* qui permet de qualifier ces espaces, notamment en termes de spécialisation économique, de fonctions de commandement politique ou culturel, ce qui peut se traduire par des degrés de hiérarchisation ;
- la *dimension relationnelle abstraite* qui permet de délimiter des espaces urbains fonctionnant comme un système cohérent, là où l'adjacence des zones bâties n'est plus le critère le plus pertinent pour regrouper des espaces urbains fragmentés.

Les périmètres urbains Ces dimensions renvoient en fait à une autre question d'importance dans la construction de modèles urbains, qui n'était considérée jusque là que de manière implicite : la question des périmètres urbains. La question du périmètre d'étude est essentielle puisque le périmètre impose les éléments qu'il est possible d'observer et de différencier au sein de la structure.

Les premiers modèles présentés, les plus anciens, décrivent des organisations monocentriques ou multipolaires à l'intérieur d'espaces urbains bornés par le périmètre de l'agglomération. Les pôles déconnectés apparaissent encore à ce stade comme des espaces complémentaires, qui ne sont généralement pas entièrement intégrés et décrits par les modèles. Mais les derniers modèles présentés, les modèles actualisés qui sont supposés représenter la ville au tournant du 21ème siècle, correspondent à des régions urbaines présentées comme des systèmes urbains. Les limites de ces régions sont par nature plus difficiles à saisir puisqu'elles peuvent intégrer plusieurs agglomérations morphologiquement distinctes, et soumises aux influences de régions voisines. La propriété de compacité, donc la fermeture de l'espace urbain, est perdue avec les phénomènes de périurbanisation et d'étalement urbain. Ces objets urbains métropolitains se présentent en fait comme des systèmes ouverts (Haggett, 1973), qui interagissent en permanence avec leur environnement.

Si la question des périmètres peut être laissée non résolue lors de la définition conceptuelle d'un modèle, elle doit forcément être résolue lors de la phase de traitement empirique des données. On en vient alors à produire des délimitations à partir de critères statistiques appliqués par exemple à la morphologie du bâti pour délimiter des agglomérations, ou à des flux de mobilité pour définir des systèmes fermés représentant des régions urbaines fonctionnelles. L'encadré 1.2 expose quelques périmètres urbains de référence qui peuvent donc être employés pour caractériser les espaces urbains, et écarter les espaces non urbanisés. Ces périmètres auront une influence sur les caractérisations et les organisations observées et donc sur les modèles qui peuvent être construits. Comme le notent Guérois et Paulus (2002) en conclusion de leur article, les délimitations morphologiques en agglomérations gardent leur intérêt pour modéliser des effets de hiérarchisation et de spécialisation économique. Elles ont cependant tout intérêt à être complétées par des délimitations fonctionnelles de type aire urbaine,

au sens de l'INSEE, pour saisir les processus d'étalement urbain qui transforment les structures urbaines actuelles.

Encadré 1.2 (*Les périmètres urbains de référence*)

Critères administratifs En France l'unité spatiale de base utilisée pour constituer des zonages urbains est très souvent la commune. Ces unités administratives qui servaient historiquement de périmètre pour déterminer les limites officielles de la ville ont pu jusqu'à un certain point être modifiées et agrandies afin d'intégrer dans une même zone les espaces urbanisés en continu. Mais ces périmètres sont aujourd'hui largement dépassés par l'étalement et la complexité des espaces urbains, qui ont mené à la définition de multiples périmètres définis selon des critères morphologiques ou fonctionnels.

Critères morphologiques Dès 1954, l'INSEE introduit le zonage en **unités urbaines** pour définir le cadre spatial des agglomérations qui s'étendent maintenant sur plusieurs communes adjacentes à une commune centrale. Actualisé périodiquement depuis, ce zonage identifie des agglomérations urbaines qui correspondent à une ou plusieurs communes adjacentes qui regroupent au moins 2000 habitants et sur lesquelles le bâti s'étend sans discontinuité de plus de 200 mètres entre deux constructions (en tenant compte des coupures telles que les cours d'eau). D'autres périmètres se basent sur l'utilisation de données sur l'occupation des sols pour identifier l'emprise des taches urbaines, selon le même critère de continuité du bâti mais avec des seuils différents de population. Par exemple, les **UMZ** (*Urban Morphological Zones*), définies pour le compte de l'agence européenne pour l'environnement (Simon et Leal, 2011), ou la base de données Europolis du projet **Geopolis** (Moriconi-Ebrard, 1994).

Critères fonctionnels Après l'introduction d'un premier zonage fonctionnel, les ZPIU (Zones de Peuplement Industriel et Urbain) créées dans les années 1960, l'INSEE a complété ses zonages urbains en 1997 avec le zonage en **aires urbaines**. Construits pour intégrer les phénomènes d'étalement urbain et de périurbanisation, ces périmètres se composent d'un pôle urbain (une unité urbaine) de plus de 5000 emplois et d'une couronne périurbaine, c'est-à-dire des communes ou unités urbaines qui envoient au moins 40% de leurs résidents travailler dans le pôle ou dans les communes attirées par le pôle. Cette approche fonctionnelle des espaces urbains où les navettes domicile-travail jouent un rôle déterminant est aussi employée pour construire des périmètres urbains officiels harmonisés au niveau européen maintenant connus sous nom de **zones urbaines fonctionnelles** (OCDE, 2013), qui correspondent aux anciennes LUZ ou aires métropolitaines.

Les échelles Les variations sur ces deux thèmes entraînent déjà une diversité dans les structures urbaines, qui augmente encore lorsque l'on s'attarde sur une dernière di-

mension qui est l'échelle. Souvent absente des modèles schématiques ou économiques, ou parfois considérée comme fixe pour des questions de simplification, la variation d'échelle est pourtant essentielle pour l'appréhension des structures urbaines. L'échelle intervient à travers la taille des unités spatiales minimales qui composent le système urbain. Ces unités spatiales donneront une vision plus ou moins fragmentée des objets urbains qui est appréhendée par les modèles de villes fractales. Elle renvoie à la définition de la ville comme un système intégré dans un système de villes (Berry, 1964), où les éléments (hommes, constructions, activités, etc.) peuvent se composer en autant de sous-systèmes qu'il y a d'échelles (Beaujeu-Garnier, 1995). L'échelle est une caractéristique clé des modèles urbains, puisque le changement d'échelle d'observation des formes urbaines peut conduire à une interprétation différente de la structure urbaine. Un système urbain polycentrique, composé d'un centre principal et de sous-centres à proximité, peut être vu à une échelle supérieure comme un système monocentrique. La centralité est alors assurée par un unique centre élargi qui à cette échelle retrouve le même pouvoir polarisateur sous forme de système gravitaire.

La ville évolue et les modèles ont suivi ses évolutions (émergence de centres multiples, spécialisation et sectorisation de l'espace, étalement des hommes et des activités, modification des relations de hiérarchisation). La majorité des modèles cités mettent en avant une dynamique urbaine interne ou une dynamique englobant l'urbanisation pour expliquer l'aboutissement des organisations observées. La dimension temporelle de la morphogenèse ne peut donc être oubliée pour saisir la diversité de ces structures. L'histoire urbaine qui a joué sur la construction des semis de villes plus ou moins anciennes et caractérisées par des degrés différents de densité et de proximité, joue aujourd'hui encore sur les espaces urbains (Bretagnolle et al., 2009).

Cinq clés de lecture semblent donc essentielles pour qualifier des éléments de structure : l'échelle, la forme, la fonction, la hiérarchie et les relations. Ces caractéristiques permettent, selon nous, d'identifier des éléments stables d'organisation, dans le sens où, quand elles sont observées, elles permettent de rattacher les organisations à une dimension temporelle, ce qui valide leur dimension structurelle.

1.2 Proposition pour un modèle simple de structure urbaine

Les modèles théoriques qui ont été présentés précédemment fournissent un cadre conceptuel de modélisation des organisations urbaines reposant sur une vision explicative de plusieurs processus de différenciation des localisations intra et inter urbaines. Observés d'un point de vue social et économique, ces processus informent sur la localisation des hommes et des activités et sur les relations qu'ils entretiennent. Ces modèles définissent une structure spatiale, que l'on peut associer à une structure physique,

même si elle n'est plus définie par un critère de continuité ou de compacité. L'étude et la définition de cette structure spatiale et physique sont indispensables pour comprendre, expliquer et planifier une forme urbaine qui se définit maintenant comme un système métropolitain changeant rapidement et marqué par de nombreux mouvements.

Notre objectif est alors de proposer un modèle qui synthétise les différentes dimensions clés évoquées précédemment afin de présenter un support pour comprendre ces mouvements inscrits dans différentes échelles et temporalités. Nous faisons l'hypothèse qu'une modélisation du cadre physique, c'est-à-dire les traces bâties laissées par les hommes, permettra d'améliorer la connaissance de mouvements inscrits dans deux temporalités différentes :

- les mouvements quotidiens correspondant par exemple à des pratiques de déplacement des individus ;
- les mouvements inscrits dans la longue durée correspondant à la morphogenèse du système.

Nous faisons l'hypothèse que la représentation méthodique d'une structure morphologique offre une source d'information pour comprendre et analyser ces différents mouvements. Cette approche introduite par les travaux pionniers de Webber (1996), reprise dans divers travaux de morphologie urbaine plus actuels (Badariotti, 2005; Sahraoui et al., 2014; Tran et al., 2014) repose sur l'idée que la compréhension de la structure spatiale des métropoles ne peut se faire que par l'articulation de trois composantes fondamentales et entremêlées : la composante d'interaction, la composante physique (canaux d'infrastructure et espaces adaptés), et la composante d'activité. Dans ce contexte la composante physique sur laquelle nous nous focaliserons apparaît déterminante puisque « (...) la fonction suprême de la configuration physique est d'accueillir les différents types d'activités qui s'y déroulent et de se prêter aux interactions entre les individus et les groupes impliqués dans ces activités. » (Webber, 1996, p. 54)

La structure urbaine sur laquelle nous nous attarderons sera alors le cadre de compréhension d'espaces urbains fonctionnant aujourd'hui comme des systèmes urbains marqués par la dissociation des lieux de résidence, de travail, de consommation ou encore de loisir. Ce sont ces éléments généraux qui une fois assemblés permettront de révéler les particularités d'organisation de chaque espace urbain, considéré ici uniquement sur le territoire français. Elle sera donc un support de connaissance et de compréhension de certaines composantes des espaces urbains qui trouvent leur intérêt pour des questions d'aménagement urbain car les questions de formes, de consommation d'espace ou de pratiques de transport sont devenues des problématiques clés pour la mise en place d'une politique de développement durable.

Afin de proposer nos objectifs de modélisation, nous introduisons deux notions de description des espaces urbains, la densité et les tissus urbains, qui seront utilisées pour

constituer un cadre d'observation des structures urbaines. Ces notions nous permettront de construire un modèle défini selon des critères de différenciation, répondant à une articulation entre concepts théoriques et données observées, qui sera exposé en fin de section.

1.2.1 La densité

La notion de densité qui exprime le plus simplement le phénomène d'accumulation « de beaucoup d'éléments en peu de la place » est intrinsèquement attachée à l'espace et à la morphologie urbaine caractérisée par une accumulation de personnes, de richesses ou encore de bâtiments. Elle permet d'attacher une mesure à un espace donné et par là, elle permet de faciliter la caractérisation et la comparaison des territoires. Cette caractérisation peut toutefois être multiple et source de confusion, puisque la densité rassemble en fait à une diversité d'indicateurs liés à la diversité des données disponibles et à la diversité des surfaces considérées. Nous renvoyons au chapitre 2 pour plus de détails quant aux divers indicateurs de densité qu'il est possible de calculer et d'étudier.

Largement employée pour décrire les territoires urbains, la densité est une notion reprise dans de nombreuses études pour quantifier les différences à l'intérieur des formes urbaines. Particulièrement mise en avant ces dernières années, elle se trouve au cœur des préoccupations actuelles en matière d'aménagement urbain. Elle intervient principalement selon deux dimensions d'études reliées à la mise en place de politiques de développement durable pour modérer la consommation d'espace et gérer efficacement les pratiques de mobilité.

Le premier objectif mobilise la notion dans le sens où la densification est souvent présentée par les aménageurs et décideurs comme un moyen à opposer à l'étalement et à la fragmentation urbaine. Dans des problématiques de renouvellement urbain et d'intensification¹ de la ville, la densité est souvent considérée comme l'outil de caractérisation et d'analyse des espaces urbains. Dans ce cadre, diverses études se sont attachées à étudier les liens entre densité et consommation d'espace, notamment pour identifier les sources d'augmentation de l'artificialisation (Tourneux, 2006), pour identifier des schémas de développement urbain (Bessy-Pietri, 2000; Guérois et Pumain, 2008), ou pour caractériser des types de formes urbaines selon leur consommation d'espace (Fouchier, 1997).

Le second objectif fait appel à la densité pour comprendre les différentes pratiques de déplacement. En premier lieu, elle constitue une composante clé pour la mise en place de politiques de transport puisqu'elle permet de discriminer des zones selon des volumes d'émission et de réception de flux. Ces zones traduisent la réalisation d'une

1. contrairement à la densification, l'intensification renvoie à l'idée d'accueillir plus d'hommes, d'activités ou de fonctions dans une enveloppe bâti fixe et préexistante (Charmes, 2010)

demande de transport selon le type d'offre de transport disponible. La densité permet donc d'une part de mesurer des potentiels d'utilisation des transports en commun, et d'autre part d'identifier les axes (routiers ou ferroviaires) sujets à la congestion en dépit de leur performance, qui affecteront le fonctionnement de l'ensemble d'un système urbain (Bavoux et al., 2005).

De plus, bien que diverses variables socio-démographiques contribuent à expliquer les différences de comportements en matière de déplacement, de nombreuses études ont montré que l'organisation et la différenciation du cadre bâti sont corrélées avec l'organisation de la mobilité. L'étude très populaire de Newman et Kenworthy (1989), interrogeant la répartition des modes de déplacement selon les configurations urbaine montre par exemple l'existence d'un lien entre densité et consommation énergétique des habitants dans 37 métropoles du monde. Dans leur étude portant sur un ensemble de quartiers de la Baie de San Francisco, (Cervero et Kockelman, 1997) montrent que des quartiers denses, possédant diverses fonctions (résidentielles et commerciales) et dont les aménagements urbains sont orientés vers les piétons auront tendance à diminuer le nombre de déplacements par automobile et à favoriser les autres modes de déplacement. Plus récemment, dans une étude sur Copenhague, (Naess, 2006) montre lui aussi l'influence de la densité, en plus de la distance au centre de Copenhague, sur les modes de déplacement. La densité de résidents et d'emplois favorise les déplacements non motorisés en semaine et augmente les distances parcourues à pied ou à vélo, tandis que la distance au centre de Copenhague augmente le nombre de déplacements, lesquels sont majoritairement effectués en voiture ou train au détriment de la marche ou du vélo. Au niveau de la France métropolitaine, (Hubert et Delisle, 2010) ont utilisé des déciles de population par densité croissante de la commune de résidence, pour analyser les comportements de mobilité. Ils observent une relation quasiment linéaire entre la vitesse des déplacements et la densité (population + emplois). En outre, l'utilisation de l'automobile décroît quand la densité augmente mais seulement à partir de la densité médiane, car l'utilisation des autres modes dépend aussi de la position et de l'extension des zones denses dans l'ensemble d'une agglomération.

Comme le montrent ces diverses études, les variations de densité pour différents indicateurs typiques (population résidente, emplois dans différents secteurs, bâtiments, etc.) sont pertinentes pour différencier des pratiques de mobilité. Même si des études récentes (Ewing et Cervero, 2010; Tran et al., 2014) ont montré que la densité ne constitue pas la variable la plus discriminante face à d'autres caractéristiques socio-économiques, les liens trouvés dans ces diverses études semblent indiquer que la densité est une variable intermédiaire, un proxy, capable d'exprimer les effets de plusieurs caractéristiques structurelles.

Si les différences de densité considérées à petite échelle peuvent apporter une information utile pour la mise en place d'une politique de transport (demande de déplacement, sources de congestion, etc.), il faut rappeler qu'elle n'a pas de valeur en soi pour

des politiques d'aménagement ou pour comprendre des motifs de déplacement. En tant que mesure, elle peut être source de confusion à cause d'effets d'agrégation masquant l'hétérogénéité de l'espace, mais aussi puisque des formes bâties variées peuvent être associées à une même densité selon la mesure considérée. En tant que notion, elle ne prend sens que lorsqu'elle est remise dans le contexte relationnel et fonctionnel des formes que l'on peut rattacher à des fonctions de centralité et d'activité qui expliqueront la densité générée par l'attractivité de certains lieux. La densité nous servira donc de proxy pour qualifier et différencier des formes du bâti, en supposant qu'il existe des liens entre la densité et certains types de formes.

1.2.2 Les tissus urbains

Plus qualitative que la densité, la notion de tissu urbain (ou *urban pattern*) fournit un moyen de description de la morphologie qui permet de considérer d'autres aspects des organisations urbaines à confronter aux problématiques d'aménagement et de pratiques de déplacement. La notion de tissu urbain offre en effet le moyen d'opérer une différenciation des formes intra-urbaines selon des caractéristiques de forme et d'usage importantes pour qualifier l'organisation des structures urbaines.

Les tissus urbains se définissent par le « cumul d'aménagements de l'espace et leurs relations les uns avec les autres » (Gauthiez, 2003), formant des ensembles homogènes ou hétérogènes issus de processus complexes combinant actions individuelles et collectives. Les tissus urbains peuvent être abordés selon différents éléments de l'organisation morphologique de la ville : la trame viaire, le parcellaire, le tissu constructif et les espaces ouverts pris en complément, ou encore l'utilisation des sols (Allain, 2004). Autant de composantes qui sont déterminantes pour la compréhension des pratiques de mobilité (Stead et Marshall, 2001), ou des logiques de structuration sociale et des phénomènes de ségrégation (Hillier et Vaughan, 2007), ou des polarisations économiques et fonctionnelles (Burger et Meijers, 2012).

Cette notion présente un intérêt dans le sens où les tissus urbains représentent le lien entre la forme physique, la forme sociale et la forme relationnelle. Qualifiés par un agencement physique relatif aux infrastructures ou aux bâtiments, les tissus urbains déterminent et sont déterminés par les fonctions et les relations qu'ils peuvent supporter. Associés à une forte inertie et stabilité, ils conditionnent les choix de localisation des activités et des individus. Ils peuvent également jouer un rôle sur les mobilités avec les possibilités et capacités de déplacement et de stationnement qu'ils offrent (Webber, 1996).

Le plus souvent, les tissus urbains sont étudiés par la caractérisation de la trame viaire ou de l'agencement du bâti, qui sont les éléments d'organisation les plus stables dans une composition morphologique où les fonctions et les statuts sociaux peuvent

changer plus rapidement. Dans l'analyse des tissus urbains sous l'angle des formes bâties, on retrouve des études intégrant des clés de description tirées de l'écologie du paysage, et portant par exemple sur la mixité d'utilisation des sols (Hess et al., 2001), ou sur la classification des bâtiments (Steiniger et al., 2008; Puissant et al., 2011). Dans ce cadre, nombreuses sont les études qui intègrent aussi une approche fractale pour caractériser les tissus urbains (Longley et Mesev, 2000; Badaritti, 2005; De Keersmaecker et al., 2004). L'angle bâti présente l'avantage de permettre de qualifier les tissus urbains selon des caractéristiques pertinentes pour des problématiques de transport, mais aussi selon des caractéristiques sociales et économiques qui élargissent le champ d'étude des organisations urbaines. Il faut noter que la trame viaire et le bâti représentent deux angles complémentaires de caractérisation des tissus urbains liés par les influences réciproques que jouent ces deux composantes sur le développement urbain. Si Tran et al. (2014) constatent que la caractérisation des formes bâties est plus pertinente que celle du réseau physique pour appréhender la mobilité, c'est sans doute car les formes bâties sont plus englobantes. On peut leur rattacher des caractéristiques liées aux types de bâtiments, à la population et aux activités présentes dans ces bâtiments. Mais par complémentarité, on peut déduire des caractéristiques liées aux réseaux qui les desservent et aux autres types d'espaces vides, notamment via l'agencement des bâtiments.

Pour répondre à nos objectifs de modélisation, l'intérêt de la caractérisation des tissus urbains sous l'angle des formes bâties est aussi qu'ils peuvent être identifiés selon un ensemble de mesures de densité. On distinguera alors deux volets d'études (Fouchier, 1997) : la densité de contenu associée à la caractérisation des populations et des activités, et la densité de contenant associée à la caractérisation du cadre bâti. En combinant ces deux volets d'étude il sera alors possible de proposer une caractérisation des formes de tissus urbains observés dans un espace de résidences, de services et d'emplois. Fonctions et niveaux de densité seront les caractéristiques que nous étudierons dans un périmètre urbain cohérent d'un point de vue relationnel, afin de concevoir un mode de construction des structures urbaines capable d'intégrer différents éléments de modélisation théorique tirés du précédent état de l'art.

1.2.3 Les objectifs de modélisation

La dernière question qui se pose est de savoir quels sont les types de forme pertinents pour recomposer des organisations urbaines. Quels sont les rapports d'opposition les plus marquants pour reconstruire des structures urbaines définies selon les dimensions de caractérisation énumérées à l'issue de l'état de l'art de la modélisation urbaine ?

Comme évoqué précédemment, l'objectif est de formaliser une représentation selon des caractéristiques de densité pour *in fine* retrouver des tissus urbains auxquels on pourra rattacher des caractéristiques de formes et de fonctions pertinentes pour différencier des organisations urbaines. Ces tissus urbains pourront servir de support à la

compréhension de mouvements inscrits dans des temporalités longues et courtes, dans l'idée que les tissus urbains sont associables à des périodes engendrant différents types architecturaux, eux-mêmes associables à différentes densités.

Avec ces notions nous pourrions intégrer un jeu d'échelles pour passer d'une description fine des objets urbains sur des grilles régulières à une description généralisée des structures observées sur une diversité de territoires urbains. Ce jeu d'échelles correspondra à un processus de traitements emboîtés des données, qui seront exposés dans le chapitre suivant, à trois échelles : une échelle micro au niveau du pixel, une échelle meso au niveau de la région d'image, et une échelle macro au niveau de l'objet urbain métropolitain.

L'échelle du pixel sera celle des données observées et utilisées pour construire des modèles aux deux échelles suivantes. Les analyses à l'échelle du pixel permettront de constituer une base de tissus urbains par regroupement de pixels associés à des densités et compositions homogènes. On cherchera alors différents rapports d'opposition caractérisant des formes bâties définies selon des critères de hiérarchisation, de forme, de fonction et de relation. Les analyses à l'échelle de la région d'image permettront de constituer une seconde base de représentation morphologique, toujours définie selon des rapports d'opposition définissant cette fois des morphotypes. Le changement d'échelle force le changement d'interprétation des organisations : les relations spatiales constitueront la caractéristique clé pour définir des structures morphologiques. Ces structures ne seront pas associées à un modèle type d'organisation, mais à une diversité d'organisations résultant des particularités des territoires qui amènent une part de variabilité dans la combinaison des modèles et des processus d'organisation.

Le tableau 1.1 présente les rapports d'opposition que nous nous proposons de modéliser dans la suite de ce travail. Ils permettront de construire une caractérisation de la structure spatiale qui pourra servir de contenant général, de référentiel, pour diverses études portant sur le contenu social et fonctionnel associé à la structure morphologique.

TABLE 1.1 – Caractéristiques et rapports d'opposition pour un modèle morphologique

Echelle	Meso (tissus urbains)	<i>vs</i>	Macro (morphotypes)
Hierarchie	Centre(s) <i>vs</i> Périphérie		Centre(s) <i>vs</i> Périphérie
Fonction	Résidentiel <i>vs</i> Activité		Résidentiel <i>vs</i> Activité
Forme	Habitat individuel <i>vs</i> collectif		Dense <i>vs</i> Peu dense
Relation	Habitat groupé <i>vs</i> dispersé		Connecté <i>vs</i> Non connecté

Cette caractérisation, nous proposons de l'étudier au sein des périmètres des aires urbaines, que l'on considère représentatif des étendues des formes urbaines métropolitaines françaises. Déterminés selon les mêmes critères sur l'ensemble du territoire français, ils sont suffisamment englobants pour contenir les différentes caractéristiques

morphologiques que nous souhaitons identifier. De plus, ces périmètres définis selon des critères fonctionnels offrent le moyens d'étudier des organisations morphologiques que nous savons cohérentes pour au moins une composante d'interaction, celle des déplacements domicile-travail. Même si l'accent est mis sur la morphologie, nous pourrions proposer un modèle d'organisation élaboré en tenant compte au moins de manière indirecte des dimensions fondamentales pour caractériser la structure urbaine.

Conclusion

Dans ce chapitre nous avons introduit les clés de description et de modélisation des espaces urbains que nous utiliserons tout au long de ce travail. En se basant sur un état de l'art des modèles d'organisation urbaine, nous avons dégagé une synthèse de cinq critères de caractérisation qui nous semblent essentiels pour proposer une représentation simple des structures urbaines du cadre bâti : l'échelle, la forme, la fonction, la hiérarchie et les relations. En utilisant ces critères, nous avons proposé des éléments conceptuels pour construire un modèle morphologique des structures urbaines qui seront observées à partir de données sur grille régulière.

Notre objectif est de proposer une représentation de l'organisation morphologique d'espaces urbains inscrits dans des logiques de fonctionnement systémique. Les mouvements et les relations entre des espaces aux fonctions différenciées définissent une structure liant la morphologie physique et sociale des espaces urbains. Les critères de définition de notre modèle reposent donc sur la création d'un référentiel qui apporte une connaissance articulée autour de la compréhension des organisations fonctionnelles observées sous l'angle des déplacements, et de la morphogenèse.

L'ambition de ce travail est de proposer une modélisation qui puisse être appliquée à n'importe quel territoire urbain français caractérisé à partir de données sur grille régulière. Cet objectif qui a nourri la définition du modèle doit se traduire dans la suite par la sélection de données adaptées pour différencier les différentes caractéristiques d'organisation données en fin de chapitre. Les éléments de modélisation et de description donnés dans ce chapitre vont donc constituer des guides à la lecture des données. L'idée sera de voir comment ces éléments vont orienter le développement de structures urbaines, toutes particulières.

Chapitre 2

Quantifier les espaces urbains : construction et choix des variables d'intérêt

Sommaire

Introduction	56
2.1 Les critères de description : quelles variables d'étude ?	57
2.1.1 Une lecture morphologique des objets urbains	57
2.1.2 La construction d'indicateurs sur les grilles régulières	65
2.1.3 Discussion	72
2.2 Le choix d'une nomenclature de tissus urbains	75
2.2.1 Modéliser l'organisation urbaine par une structure radioconcentrique	75
2.2.2 Classification des tissus urbains par règles de décision	79
2.2.3 Discussion	87
Conclusion	88

Introduction

Ce chapitre aborde le problème de la représentation des formes urbaines métropolitaines par la question des données à utiliser pour les extraire. L'enjeu est ici de définir les thématiques de données et les mesures associées, nécessaires pour identifier les éléments des modèles théoriques présentés dans le chapitre précédent.

Nous cherchons donc dans ce chapitre à identifier un ensemble de variables définissant la première échelle d'observation à partir de laquelle il sera possible de réaliser des tâches de classification. Ces variables doivent nous permettre de construire une image des formes urbaines. L'image est l'objet qui nous intéresse au sens de la représentation abstraite déterminée par ce qui est observé et simplifié : la morphologie. Mais c'est aussi l'objet qui nous intéresse au sens du tableau matriciel composé de pixels qui seront classés par des méthodes de traitement d'image. Le travail sur les grilles régulières assimilables à des images, que nous traiterons sous forme multivariée, a donc pour but d'identifier dans un premier temps différents types de tissus urbains. Ces types déterminés par les principes d'organisation et d'opposition identifiés dans le chapitre 1 se combinent en une première image des organisations urbaines observées à l'échelle meso-géographique.

Nous traiterons dans un premier temps le choix des variables d'étude, et nous verrons que différentes mesures de densité fournissent un cadre d'analyse pour révéler les différences de configuration morphologique. Thème d'intérêt pour appréhender différentes facettes de l'urbain, liées au transport ou à l'aménagement, la densité représente un concept largement employé qui peut se décliner en de multiples mesures liées à la population et au cadre bâti. Nous détaillerons les sources disponibles et un ensemble d'indicateurs qu'il est possible de construire sur des grilles régulières.

Puis, à partir d'un travail préparatoire sur quelques agglomérations tests, nous verrons qu'il est possible de construire une nomenclature de tissus urbains caractérisant des rapports d'oppositions morphologiques. Pour cela nous considérerons deux types d'*a priori* sur les données : un *a priori* thématique et un *a priori* morphologique. Nous désignons par le terme *a priori* des connaissances extérieures relatives aux caractéristiques d'organisation introduites précédemment. L'utilisation de ces *a priori* traduit la volonté d'intégrer une connaissance utile pour la réalisation des classifications. À partir d'une combinaison d'indicateurs de densité de population et de bâtiments et d'une transformation des mailles du référentiel spatial des données, nous introduirons la nomenclature des tissus urbains qui servira de base à l'élaboration de notre représentation des structures urbaines.

2.1 Les critères de description : quelles variables d'étude ?

La détermination d'un référentiel pour construire la structure des formes urbaines métropolitaine passe par l'intégration d'un *a priori* thématique qui opère par le choix des variables de description. Ce choix guidera dans la suite la classification des mailles en fonction des objets qu'elles regroupent. Il doit donc être déterminé en fonction des objectifs de représentation et d'analyse de la composition des structures urbaines à partir d'unités spatiales qui doivent avoir un sens plus révélateur que les simples mailles carrées d'une grille.

Ces *a priori*, ou critères de description des espaces urbains, sont contraints par les enjeux de représentation que nous nous sommes fixés, qui détermineront en partie les données disponibles et utilisables. Les variables dont on peut disposer doivent être issues de données attachées à des unités spatiales plus petites ou égales aux unités spatiales des grilles régulières à mailles de 200 mètres de côté, ou de données faciles à désagréger sur les grilles régulières. Cette contrainte nous a poussé à arbitrer en faveur de la morphologie, en supposant que cette dimension d'étude offrait suffisamment d'information pour construire référentiel d'étude. Ce référentiel se veut simple mais suffisamment robuste et pertinent pour définir un support d'étude qui puisse dans le futur être enrichi par d'autres données thématiques.

En se basant sur le référentiel spatial des grilles régulières, nous introduisons ici les premières bases de cette représentation au format image reposant sur l'utilisation de mesures de densité de population et de bâtiments. Nous détaillerons ensuite la construction d'indicateurs sur ces grilles à partir de données désagrégées issues de la BD Topo® de l'IGN.

2.1.1 Une lecture morphologique des objets urbains

Le problème de la représentation d'objets urbains est considéré sous l'angle de la morphologie urbaine et plus particulièrement sous l'angle de la densité. La densité renvoie en fait à une diversité d'indicateurs liée à la diversité des données disponibles, associées aussi bien à des densités de contenant qu'à des densités de contenu, mais aussi à la diversité des surfaces considérées. L'utilisation des grilles régulières nous permet de lever une première source de confusion : toutes les données présentées ici seront attachées à la même unité spatiale de référence qui sera la maille de 200 mètres de côté d'une grille régulière. Nous verrons toutefois que l'utilisation de plusieurs sources de données sera une condition nécessaire pour fournir une caractérisation aussi complète que possible de la morphologie des espaces urbains. En effet l'association des thèmes de densité de contenant et de contenu permet de détecter certaines modulations mor-

phologiques et fonctionnelles de densité qui ne seraient pas observables si seul un des thèmes était considéré (Fouchier, 1997).

2.1.1.1 La population : base de repérage des établissements humains

La mesure de la population et de la densité de population constituent une base de repérage des établissements humains employée dans les plus anciennes définitions statistiques de l'urbain. Sa déclinaison la plus simple, le nombre d'habitants divisé par la surface des unités spatiales de mesures tirées par exemple des fichiers du recensement, reste un thème largement employé comme source de différenciation des formes d'établissement humain qualifiées selon un éventail de formes hiérarchisées par des niveaux de densité.

La population résidente, premier thème à être diffusé sur des données carroyées au niveau européen et français, constitue une source d'information pertinente et déjà exploitée dans diverses études. Elle est par exemple employée par l'OCDE pour identifier des centres urbains morphologiques (Dijkstra et Poelman, 2012). Elle est employée par la commission européenne pour définir une nomenclature harmonisée des espaces urbains et ruraux (Eurostat, 2010), et reprise au niveau français pour construire le zonage en bassin de vie de 2012 (Brutel et Levy, 2012).

En France, ces données sont diffusées par l'INSEE sur des carroyages à mailles de 1 km de côté et 200 mètres de côté (cf. figure 2.1 et encadré 2.1), suite aux préconisations de la directive européenne INSPIRE (Annoni, 2003). Les données de population sont issues des Revenus Fiscaux Localisés (RFL) de 2010 des ménages, c'est-à-dire les fichiers de revenus et d'imposition, et de taxe d'habitation. Cette source de données différente du recensement fournit une « vision partielle de la population » pour deux raisons énoncées dans la documentation disponible sur le site de l'INSEE¹ : les personnes non référencées à la taxe d'habitation ne sont pas prises en compte ; et les étudiants sont généralement localisés au domicile de leurs parents. Le processus d'affectation de la population aux mailles de 200 mètres de côté résulte d'un processus de géolocalisation en deux étapes :

1. Chaque ménage est localisé par une coordonnée géographique qui correspond à « un point intérieur - généralement proche du centre géométrique - de la parcelle cadastrale principale où réside le ménage ».
2. La population par carreaux correspond au comptage des individus des ménages localisés dans chaque carreau.

Prise en tant que variable d'étude isolée, la population fournit par exemple plusieurs

1. http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=donnees-carroyees&page=donnees-detaillees/donnees-carroyees/donnees_carroyees_geolocalisation.htm

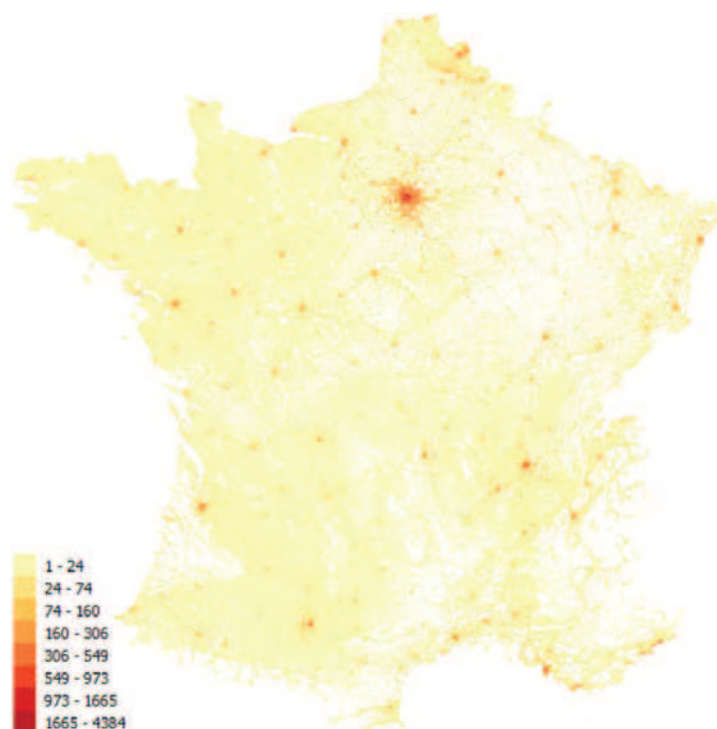


FIGURE 2.1 – Population résidente sur la grille à mailles de 200 mètres de côté de la France

dimensions d'étude et de comparaison des espaces urbanisés observés selon : leur taille ; leur densité ; l'inégalité de répartition de la population ; et le degré de *clustering* c'est-à-dire les formes de distributions spatiales (groupés ou aléatoires) des zones les plus denses (Tsai, 2005). Ces différentes dimensions renvoient à l'observation de modèles théoriques de type centre-périphérie ou gradient décroissant, en supposant que la densité peut être associée à un certain degré de centralité. Elles sont donc employées pour caractériser des formes métropolitaines, mais peuvent être étudiées aussi bien à partir de la population que d'autres mesures de densité de contenu ou de contenant.

La figure 2.2 illustre toutefois l'influence des données sur les représentations et les modèles théoriques qu'il serait possible de tester. Cette figure correspond à la population spatialisée sur des grilles à mailles de 1 km et 200 mètres de côté pour la communauté d'agglomération de Besançon. Elle révèle notamment que si l'on cherche à faire émerger une structure urbaine en se servant simplement de la population, alors on obtiendra des résultats différents selon l'échelle d'étude choisie.

Sur la grille à 1 km de résolution (figure 2.2a), la zone la plus dense coïncide globalement avec la commune de Besançon. Les communes qui gravitent autour sont associées à des densités moins importantes mais on détecte toutefois des variations locales qui permettent d'identifier les centres de presque toutes les communes. Sur la grille à 200 mètre de résolution (figure 2.2b), l'organisation de la commune centrale

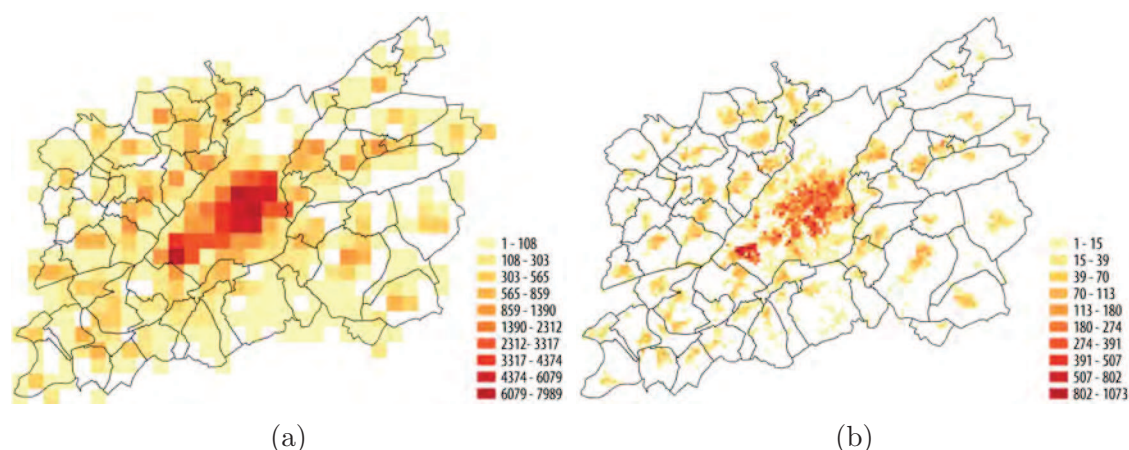


FIGURE 2.2 – Grilles de population sur la communauté d’agglomération du Grand Besançon : 2.2a Mailles d’1 km de côté, 2.2b Mailles de 200 mètres de côté

semble différente. La zone la plus dense se situe au Sud-Ouest de la commune et correspond en fait au quartier d’habitat collectif de Planoise. Le centre historique de Besançon possède, lui, une densité légèrement moins importante et se distingue à peine du reste de l’agglomération dense.

Cette différence d’organisation n’est pas sans rapport avec la question de la ventilation des présences selon le moment de la journée et à la mixité fonctionnelle des lieux (Batty et al., 2004). Nous ne disposons que de la population résidente, ce qui change les caractéristiques d’organisation, et notamment les zones les plus denses que nous pourrions détecter si la population active, ou une combinaison des deux étaient disponibles.

Sur ces grilles régulières à mailles de 200 mètres de côté, l’information beaucoup plus locale ne permet pas de construire par une simple étude univariée une représentation qui pourrait être confrontée aux modèles théoriques du chapitre 1. Elle ne permet pas à elle seule de modéliser des rapports d’opposition qui révèlent les structures en termes de hiérarchie (centre vs. périphérie) ou de fonction (résidentiel vs. activité). La finesse de localisation constitue une source d’information inédite qui doit être combinée avec d’autres thématiques de densité pour obtenir une meilleure description des organisations urbaines.

2.1.1.2 Le cadre bâti : complément pour appréhender les tissus urbains

Le cadre bâti constitue lui aussi une base d’étude largement employée qui permet d’avoir une vue différente des espaces urbains. Au cadre bâti, on associe des mesures de densité de contenant qui se déclinent en un nombre encore plus important de variables que les mesures de densités de contenu. De multiples indicateurs sont par exemple définis de manière officielle pour être utilisés comme des outils servant à la

Encadré 2.1 (*Historique de la diffusion des données carroyées de l'INSEE*)

Fin 2010 : Diffusion de la population résidente de 2007 sur des grilles à mailles de 1 km de côté. Sources : RFL, cadastre + informations sur l'occupation des sols pour les communes dont le cadastre n'était pas numérisé.

Septembre 2011 : Diffusion de la population résidente de 2009 sur des grilles à mailles de 200 mètres de côté, et actualisation de celles à 1 km. Sources : RFL, cadastre maintenant disponible sur la France entière.

Janvier 2013 : Diffusion de 23 nouvelles variables socio-économique issues du RFL 2010 sur les grilles à mailles de 200 mètres de côté. Elles concernent :

- La population résidente (totale, par classes d'âge) : 9 var.
- Les ménages (total, taille, présents depuis + 5 ans, propriétaire / locataire, en logement individuel / collectif, déclarant une résidence principale / secondaire, surface des résidences principales) : 10 var.
- Les revenus (UC, médiane du revenu par UC, ménages au dessus/en dessous du seuil de haut/bas revenu) : 4 var.

La confidentialité de certaines variables sensibles est assurée par des méthodes de permutation et d'imputation appliquées aux mailles isolées contenant un faible nombre de ménages.

Février 2013 : Suspension de la diffusion de toutes les données carroyées à 200m suite à la publication d'un article du *Canard Enchaîné* qui mettait en doute la confidentialité des données diffusées, mais sans prendre en considération les méthodes de floutage. L'INSEE décide de modifier la chaîne de production et d'abandonner la diffusion de données floutées sur la maille de 200m car des dysfonctionnements techniques ont entravé l'anonymisation d'une faible proportion de carreaux.

Novembre 2013 : Diffusion de 18 nouvelles variables socio-économiques issues du RFL 2010 :

- La population résidente (totale par carreau/rectangle, classes d'âge) : 10 var.
- Les ménages (total, taille, présents depuis + 5 ans, propriétaire, en logement collectif, surface des résidences principales) : 6 var.
- Les revenus (somme des revenus par UC, ménages en dessous du seuil de bas revenu) : 2 var.

La méthode de diffusion et de floutage des données a changé après une validation de la nouvelle méthodologie par le CNIS. A l'exception de la population résidente qui est diffusée sans floutage sur les carreaux, toutes les variables sont diffusées en zones sensibles sur des rectangles de taille variable, agrégeant suffisamment de ménages pour permettre la diffusion des données. Des traitements statistiques de type « winsorisation » ont aussi été appliqués sur certaines variables sensibles.

programmation et à la réglementation urbaine. On retrouve dans ce cadre les mesures suivantes, associées à l'utilisation de données localisées à l'échelle de la parcelle :

- SHOB : la surface de planchers calculée à partir du mur de façade extérieur ;
- SHON : la SHOB à laquelle on soustrait les surfaces hors œuvre (combles, sous-sols, balcons, etc.) ;
- COS : le rapport entre la SHON et la surface de terrain ;
- CES : le pourcentage de surface bâtie au sol (la SHOB du rez-de-chaussé).

L'analyse d'images satellites et de photographies aériennes fournit aussi une source importante d'informations pour quantifier la morphologie des espaces bâtis. Les bases d'occupation des sols dérivées de ces analyses (cf. encadré 2.2) permettent en effet de caractériser les espaces urbains à des macro et meso-échelles, selon des enveloppes bâties représentant des taches urbaines (Cuniberti et al., 2005) ou selon différentes catégories de tissus urbains. Elles offrent un référentiel spatial auquel il est possible d'associer des mesures de densité de surface bâtie, ou de surface imperméabilisée. Elles offrent en plus le moyen de dériver de multiples indicateurs permettant de quantifier les formes et organisations urbaines, souvent selon des approches et des mesures tirées de l'écologie du paysage (Turner, 2005).

Les indices les plus simples concernent les tailles des différents périmètres urbains (surfaces urbanisées en continu, en discontinu, cumul des deux), le nombre de régions composant l'espace urbain ainsi que leur taille moyenne ou variance. Il existe de nombreux indices qui fournissent une caractérisation des formes selon différents critères de compacité, de linéarité des formes, de régularité de leur bord, ou encore de dimension fractale (Herold et al., 2002; Huang et al., 2007; Schneider et Woodcock, 2008; Thomas et al., 2012). Les caractérisations des formes peuvent aussi être topologiques et donc faire appel à des variables relationnelles pour définir des indices de centralité des formes par rapport à une forme de référence (Schwarz, 2010) ou de « vertébralité » d'un réseau de transport de référence guidant fortement le développement urbain (Marshall, 2009). En plus de ces mesures morphologiques calculées dans un espace à deux dimensions, on peut aussi mentionner les indices concernant le volume urbain mesurant la verticalité des lieux (Koomen et al., 2009; Salvati et al., 2013).

Complétée par d'autres sources d'information, l'imagerie satellitaire et aérienne offre aussi le moyen de produire des bases de données beaucoup plus fines permettant une caractérisation des espaces bâtis à des micro-échelles. La **BD Topo**² de l'IGN constitue la source d'information la plus précise sur le cadre bâti. Cette base de données qui fait partie du RGE (Référentiel à Grande Echelle) contient une description de précision métrique pour la France entière. Définie pour être exploitée à des échelles allant du 1/5 000^e au 1/50 000^e, elle fournit 10 thèmes d'information sur le territoire

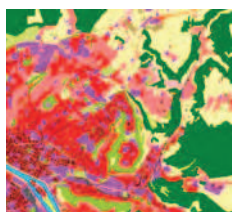
2. <http://professionnels.ign.fr/bdtopo>

Encadré 2.2 (*Les bases d'occupation des sols*)

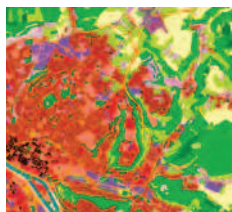
Les bases de données d'occupation des sols sont des inventaires exhaustifs de l'occupation biophysique des terres, déterminés à partir d'images satellites et/ou de photographie aérienne. Ces bases de données qui ne sont pas spécifiquement destinées à réaliser des études urbaines offrent une connaissance générale des territoires permettant notamment le suivi et la gestion du développement urbain, la mise au point d'indicateurs et l'identification de périmètres d'études. Elles se différencient selon trois critères d'utilisation : l'échelle d'utilisation, l'unité spatiale de référence, la nomenclature.



La base de données de référence à l'échelle européenne est la base de données **Corine Land Cover** (CLC), réalisée pour le compte de l'agence européenne pour l'environnement (EEA). Constituée pour répondre aux besoins nationaux et européens de suivi et de gestion des territoires, elle offre une cartographie du territoire européen à l'échelle du 1/100 000^e. Elle est composée d'unités spatiales polygonales d'au moins 25 ha de surface identifiées par photo-interprétation assistée par ordinateur (PIAO). Ces polygones correspondent à des zones homogènes couvrant l'ensemble du territoire selon une nomenclature d'occupation des sols contenant 44 postes hiérarchisés en 3 niveaux.



L'**Urban Atlas** réalisé lui aussi pour le compte de l'EEA est une base de données ciblée sur la connaissance de l'occupation et l'utilisation des sols dans les plus importantes zones urbaines européennes. Réalisée par PIAO pour une utilisation à l'échelle du 1/10 000^e, elle couvre les périmètres des LUZ (cf. chapitre 1) selon une représentation en polygones de taille supérieure à 0,25 ha. Sa nomenclature est proche de celle de CLC, mais elle est beaucoup plus détaillée pour la catégorie des territoires artificialisés, qui sont décrits par types de tissus urbains donnés en fonction de pourcentages de surface imperméabilisée.



Au niveau français, plusieurs communautés d'agglomérations, départements ou régions se dotent de leurs propres bases d'occupations des sols appelées de **MOS** (Mode d'Occupation des Sols). Construits par PIAO, ces MOS sont définis selon des nomenclatures plus spécifiques et adaptées pour représenter les particularités de chacun de ces territoires à des échelles allant du 1/5000^e au 1/50 000^e. Ces bases de données constituent des sources pertinentes d'information sur les milieux urbains qui sont dans certains cas détaillés de façon plus précise en termes d'utilisation des sols et plus particulièrement des bâtiments.

et ses infrastructures représentés sous forme ponctuelle, linéaire ou surfacique. Elle représente donc une source d'information adaptée pour compléter la quantification des espaces urbains avec différentes mesures de densité de contenant compatibles avec les grilles régulières de population à 200 mètres. Des indicateurs de densité de contenant peuvent être calculés par agrégation d'informations concernant certaines entités polygonales issues du thème Bâti. Ce thème offre une description des bâtiments distingués en trois classes :

- le **bâti industriel** qui contient les bâtiments caractérisés par une fonction industrielle, commerciale ou agricole ;
- le **bâti remarquable** qui contient les bâtiments caractérisés par une fonction administrative, religieuse, sportive ou relative au transport ;
- le **bâti indifférencié** qui contient tous les autres bâtiments, sans que leur fonction soit spécifiée. Dans la pratique, on observe que cette couche contient des bâtiments à usage résidentiel, administratif et de service. Elle contient aussi des polygones qui correspondent à des parkings couverts, des parkings individuels, ou des abris de jardin.

La représentation surfacique des bâtiments correspond à deux types de modélisation géométrique illustrés en figure 2.3. Dans certains cas, comme Bordeaux, la surface des bâtiments est désagrégée au niveau cadastral. Mais dans d'autres cas, comme Besançon, la surface correspond à l'emprise au sol du bâtiment telle qu'elle peut être détectée à partir de vues aériennes. Cette modélisation fournit donc une représentation non pas à l'échelle du bâtiment, mais à l'échelle de l'ilot dans les zones à très haute densité de bâtiments comme les centres historiques. Comme nous le verrons dans la suite, cette différence d'échelle de modélisation peut avoir un impact sur le calcul de certains indicateurs.

Chaque classe de bâtiments possède aussi une description attributaire qui offre des informations sur la nature des bâtiments (sauf pour le bâti indifférencié), sur leur hauteur, ou encore sur leurs altitudes minimales et maximales. Cette caractérisation attributaire fournit le moyen de filtrer les bâtiments pour ne sélectionner que ceux ayant des fonctions associables à un caractère urbain. A partir des bâtiments sélectionnés il sera alors possible de quantifier la densité de contenant à partir d'une diversité d'indicateurs présentés dans la section suivante.

Dans la suite de ce travail, nous nous basons donc sur deux thématiques de données, la population résidente observée sur les carroyages de l'INSEE, et le cadre bâti étudié à partir d'informations tirées de la BD Topo® de l'IGN. Le choix de se baser uniquement sur ces deux sources informant principalement sur la morphologie, mais aussi nous le verrons sur la fonction des espaces urbains, découle du cadre technique que nous avons choisi, à savoir l'utilisation de grilles régulières. Ces grilles nous permettent de

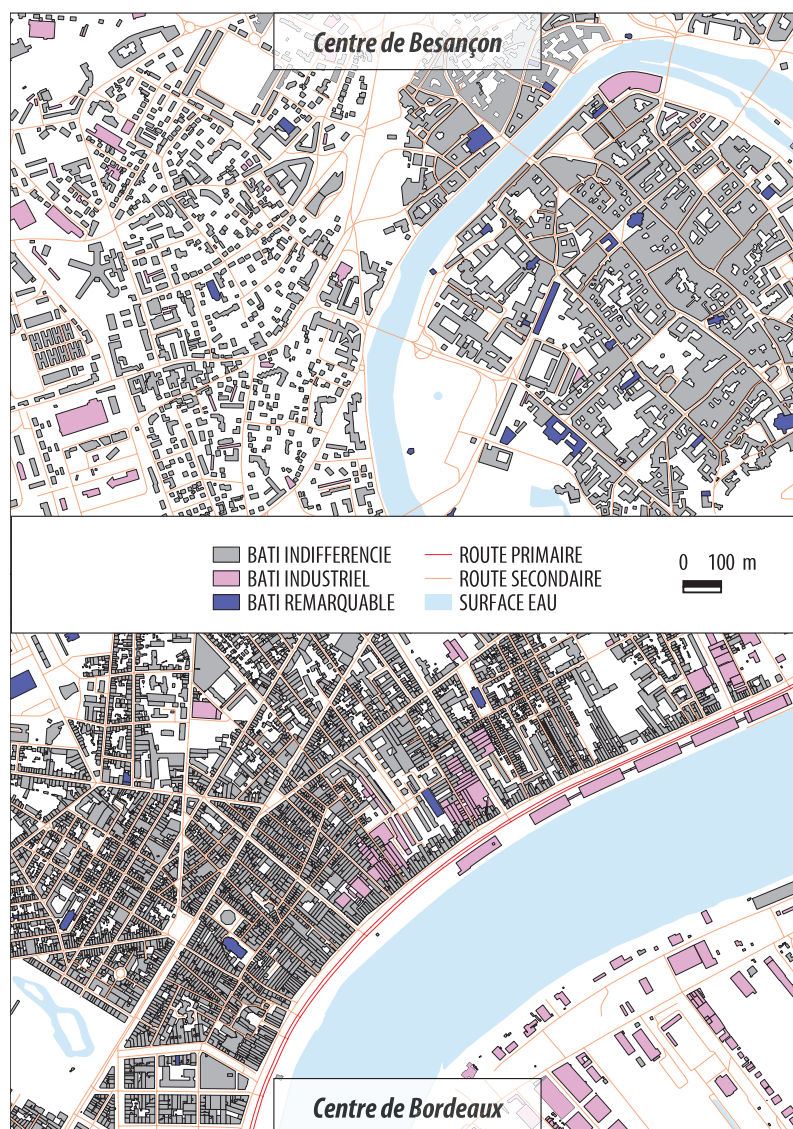


FIGURE 2.3 – Modélisations géométriques des bâtiments de la BD Topo®

construire de nouvelles images des organisations urbaines, des images fines et détaillées qui ne sont pas observables à partir de données spatialisées sur les maillages en commune ou en IRIS. Nous excluons automatiquement l'utilisation de telles données, qui étant soumises au MAUP (*Modifiable Areal Unit Problem*) tel que le définit [Openshaw \(1984\)](#), poseraient le problème du transfert sur le maillage plus fin des grilles régulières.

2.1.2 La construction d'indicateurs sur les grilles régulières

Les données dont nous disposons, à savoir la population résidente fournie par l'INSEE, sont spatialisées sur une maille de 200 mètres de côté. Un carroyage fin tel que celui-ci permet de différencier l'espace en agrégeant des informations venant de bases

de données cartographiques comme la BD Topo® de l'IGN (Lajoie, 1992). En transférant ces informations désagrégées sur ces carreaux de 200 mètres de côté, nous pouvons donc confronter la densité de la population résidente à celle du bâti et évaluer la texture du tissu bâti en fonction de la forme et de la répartition spatiale des bâtiments. L'idée est alors ici de déterminer des indicateurs de densité pour mener à bien cette caractérisation des tissus urbains. Il s'agit maintenant de sélectionner et de combiner les plus pertinents afin de caractériser au mieux ces tissus urbains.

2.1.2.1 Les indicateurs possibles de densité

Le transfert d'information de la BD Topo® sur les grilles régulières que nous traitons ici est un problème qui se résout facilement. Les polygones décrivant l'emprise des bâtiments étant majoritairement de taille inférieure aux mailles des grilles régulières, le problème de transfert se rapporte à un simple problème d'agrégation des mesures calculées à l'échelle du bâtiment. De plus, on peut supposer que les polygones qui représentent les bâtiments correspondent à des unités spatiales élémentaires et donc non modifiables (ce qui n'est pas toujours le cas, comme nous l'avons vu dans la section précédente). De là, on peut postuler que ces mesures se répartissent de façon homogène sur l'ensemble de la surface des polygones. Ces deux points permettent de définir divers indicateurs à l'échelle de la grille comme des mesures statistiques (somme, moyenne, écart-type) calculées localement après intersection spatiale de la grille et des bâtiments.

La présentation des indicateurs que nous faisons ici n'est bien entendue pas exhaustive. Nous nous focalisons sur les indicateurs qu'il est possible de calculer sans distinction des trois classes de bâtiments. De plus, nous ne présentons pas d'indicateurs issus du croisement des données de population et de bâti, puisque les deux thèmes seront ensuite intégrés dans des analyses multivariées.

Sélection des bâtiments Parmi l'ensemble des bâtiments décrits géométriquement dans la BD Topo®, tous ne présentent pas un intérêt pour élaborer des indicateurs à utiliser dans des analyses urbaines. Pour éviter de générer un bruit sur les données, nous faisons le choix de filtrer en entrée certains bâtiments dont la fonction permet de les identifier comme non pertinents pour qualifier des tissus urbains. Nous excluons des bâtiments des couches « bâti industriel » et « bâti remarquable » lorsque l'attribut de nature permet de les identifier comme spatialement et fonctionnellement hors du cadre urbain, ce qui correspond aux valeurs suivantes :

- Bâti industriel : { serre ; silo }
- Bâti remarquable : { péage ; arène ou théâtre antique ; château ; fort, blockhaus, casemate ; tour, donjon, moulin }

Ces bâtiments sont très rares dans l'ensemble des données, puisqu'une très grande majorité de bâtiments est affectée à la couche de bâti indifférencié.

Indicateurs à l'échelle du bâtiment L'intersection de la grille régulière et des polygones des bâtiments aboutit à deux configurations : soit un bâtiment est entièrement contenu dans une maille, soit il chevauche plusieurs mailles. Les indicateurs qui servent de base au calcul de tous les autres indicateurs correspondent alors aux différentes surfaces qu'il est possible de calculer :

- s_{bat} : surface d'un bâtiment entier (calculé avant l'intersection avec la grille)
- s_{morc} : surface d'un bâtiment ou morceau de bâtiment entièrement inclus dans une maille (calculé après l'intersection avec la grille)
- s_{round} : surface d'un bâtiment ou morceau de bâtiment entièrement inclus dans une maille, arrondie à 5 mètres carrés près.

Il faut rappeler qu'à l'échelle du bâtiment, nous disposons d'un indicateur de hauteur h des bâtiments figurant initialement dans la BD Topo®. Cet attribut offre la possibilité de calculer un indicateur supplémentaire, le volume v_{morc} qui fournit une information sur l'espace tridimensionnel occupé par un bâtiment :

$$v_{morc} = s_{morc} \times h \quad (2.1)$$

Enfin en complément de ces indicateurs nous introduisons un ratio r qui fournit une mesure de la portion de bâtiment contenue dans une maille :

$$r = \frac{s_{morc}}{s_{bat}} \quad (2.2)$$

Indicateurs à l'échelle de la maille A partir de ces quelques indicateurs calculés à l'échelle du bâtiment, il est possible de construire une grande diversité d'indicateurs sur les grilles régulières résultant des différentes agrégations possibles. Dans ce cadre nous distinguons deux types d'indicateurs : les indicateurs basés sur des mesures dans un voisinage d'ordre 1, composé d'une maille ; les indicateurs basés sur des mesures dans un voisinage d'ordre 2, composé d'une maille centrale et de ses 8 plus proches voisins.

Les indicateurs calculés dans un voisinage d'ordre 1, c'est-à-dire par agrégation des bâtiments b appartenant à une maille i , peuvent être assimilés à des mesures de densité de contenant. Leurs définitions respectives sont données dans le tableau 2.1.

Les indicateurs calculés dans un voisinage d'ordre 2 sont calculés par agrégation des bâtiments b se trouvant dans un voisinage N_i de 3×3 mailles centré sur une maille i . Ils peuvent être particulièrement intéressants pour mesurer la diversité ou la régularité des

TABLE 2.1 – Définitions d'indicateurs de densité de contenant calculés dans un voisinage d'ordre 1

Définition	Description
$S_{bat}(i) = \sum_{b \in i} s_{bat}(b)$	Surface totale des bâtiments intersectant une maille
$S_{morc}(i) = \sum_{b \in i} s_{morc}(b)$	Surface totale des bâtiments contenus dans une maille
$NB_{bat}(i) = \#\{b \in i\}$	Nombre de bâtiments intersectant une maille
$NB_{bat}(i) = \sum_{b \in i} r(b)$	Nombre de bâtiments pondéré par la surface contenue dans une maille
$S_{moy}(i) = \frac{\sum_{b \in i} s_{morc}(b)}{NB_{bat}(i)}$	Surface moyenne des bâtiments contenus dans une maille
$V(i) = \sum_{b \in i} v_{morc}(b)$	Volume total des bâtiments contenus dans une maille
$H_{moy}(i) = \frac{\sum_{b \in i} h(b)}{NB_{bat}(i)}$	Hauteur moyenne des bâtiments intersectant une maille
$STD_{morc}(i) =$	Écart-type de la surface des bâtiments contenus dans une maille
$\sqrt{\frac{1}{NB_{bat}(i)} \sum_{b \in i} [s_{morc}(b) - S_{moy}(i)]^2}$	

formes des bâtiments. Ils présentent l'intérêt de fournir un complément d'information à la densité qui peut parfois fournir une fausse image d'homogénéité. Ils rappellent que ce processus d'agrégation peut masquer des formes bâties présentant une grande variabilité.

Un moyen simple d'obtenir une telle mesure de diversité est par exemple le calcul d'un écart-type dans un voisinage de 3×3 mailles :

$$STD_{3 \times 3}(i) = \sqrt{\frac{1}{NB_{bat}(N_i)} \sum_{b \in N_i} [s_{morc}(b) - S_{moy}(N_i)]^2} \quad (2.3)$$

Un grand nombre d'indicateurs pourraient être introduits en complément de cette mesure d'écart-type. Nous n'introduirons en complément qu'une mesure d'entropie pouvant servir de mesure de « désordre », permettant d'observer l'hétérogénéité et la complexité de l'organisation. Largement employée en écologie du paysage et dans diverses études urbaines (Thomas, 1981; McGarigal et al., 2002), l'entropie constitue une mesure de diversité assez standard. Elle est à mettre en relation avec les mesures de dimensions fractales qui renvoient, elles, à une forme différente mais proche de mesure de la complexité.

L'entropie est en pratique calculable de différentes façons. Nous retiendrons ici une mesure locale basée sur l'entropie de Shannon, similaire à celles qui peuvent être employées en traitement d'image (Sonka et al., 2008; Yan et al., 2003). Elle est définie localement pour les mailles i d'une grille à partir d'une collection de variables aléatoires discrètes X_i à valeurs dans $\mathcal{X}_i = \{x_{i1}, \dots, x_{im}\}$. Cet ensemble correspond aux m valeurs possibles des surfaces discrétisées s_{round} observées dans un voisinage N_i de 3×3 mailles. Dans ce voisinage, la probabilité P_{ij} que la surface d'un bâtiment soit $X_i = x_{ij}$ peut être estimée par :

$$P_{ij} = \frac{n_j}{NB_{bat}(N_i)} \quad (2.4)$$

où n_j est le nombre de bâtiments dont la surface s_{round} vaut x_{ij} . L'entropie locale de la surface des bâtiments est alors donnée par :

$$ENT_{3 \times 3}(i) = - \sum_{j=1}^m P_{ij} \log P_{ij} \quad (2.5)$$

La figure 2.4 illustre les variations selon des configurations typiques de bâtiments de l'entropie calculée dans ces voisinages 3×3 mailles : plus les formes (qualifiées par les surfaces arrondies) sont régulières, plus l'entropie est basse.

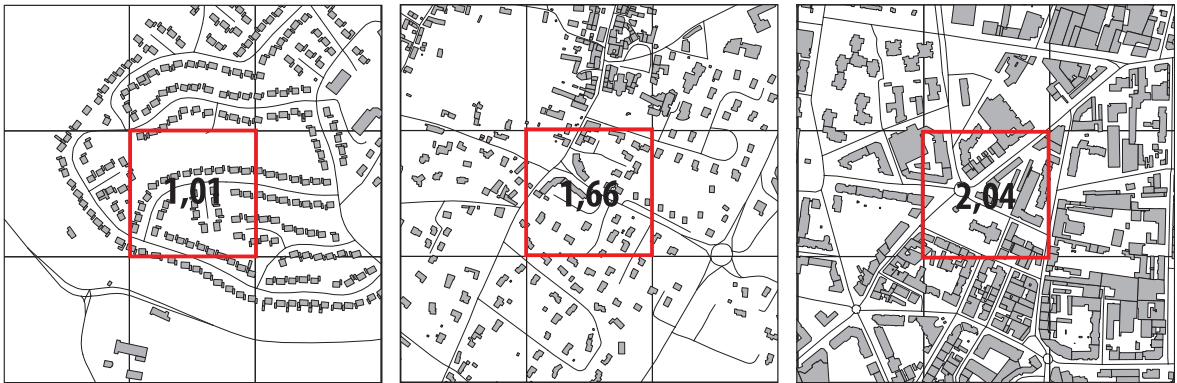


FIGURE 2.4 – Variations de l'entropie selon les configurations de quartiers observées sur l'aire urbaine Lyonnaise

Indicateurs complémentaires Les indicateurs présentés jusqu'ici correspondent à l'agrégation d'informations relatives au cadre bâti. La BD Topo® offre toutefois la possibilité de constituer des indicateurs relatifs à d'autres types d'occupation des sols en milieu urbain. Nous introduisons ici deux indicateurs complémentaires pour comprendre les liens entre espaces bâtis : la densité des réseaux viaires et des cours d'eau. Ces types d'occupation du sol représentent une sorte de squelette qui peut avoir un rôle connecteur ou au contraire séparateur entre des zones urbanisées qui sont morphologiquement distinctes.

Pour les réseaux viaires, comme pour le cadre bâti, il est possible de définir une variété d'indicateurs de densité qui pourraient en plus servir à délimiter des espaces urbains (Borruso, 2003). Toutefois nous ne nous intéressons ici qu'à leur complémentarité à l'espace bâti. Dans ce cadre nous considérons simplement un indicateur d'emprise au sol de ces réseaux en termes de surface occupée. Le calcul de cet indicateur repose sur le découpage des r tronçons du réseau routier extraits de la BD Topo® par intersection avec la grille. Puisque ces tronçons possèdent une représentation linéaire, pour obtenir une mesure de surface de réseau par maille i , il faut utiliser l'attribut de largeur disponible initialement dans la BD Topo® :

$$INFRA(i) = \sum_{r \in i} \text{longueur}(r) \times \text{largeur}(r) \quad (2.6)$$

Le calcul de l'indicateur pour les cours d'eau se fait selon le même processus que pour les surfaces bâties, mais à partir de la BD Carthage®³ (plus précise pour ce thème que la BD Topo®) qui fournit une représentation des surfaces en eau caractérisées par des attributs précisant leur nature (cours d'eau, lac, etc.).

Pour caractériser des tissus urbains sur les grilles régulières, nous disposons à ce stade d'une base de données morphologiques (cf. encadré 2.3) fournissant divers indicateurs relatifs à la population et aux surfaces bâties pour la France entière. Si la diversité de ces indicateurs peut être un atout pour caractériser plus précisément les tissus urbains, il semble toutefois peu adéquat de chercher à tous les exploiter simultanément. Nous verrons dans la suite qu'il vaut mieux réduire cet ensemble d'indicateurs et ne sélectionner que les plus pertinents par rapport à la tâche envisagée.

Encadré 2.3 (*Base de données morphologiques*)

Couverture : France entière décomposée en aires urbaines, régions et départements
Géométrie : Grilles à mailles carrées de 200 mètres de côté, Projection LAEA (EPSG : 3035)

Attributs : 14 champs correspondant à

- 1 indicateur de densité de contenu : la population résidente
- 11 indicateurs de densité/diversité de contenant bâti
- 2 indicateurs de densité d'infrastructure

Sources de données : INSEE, IGN (BD Topo®, BD Carthage®, GEOFLA®)

SGBD : PostgreSQL/PostGIS (routines d'automatisation des calculs en PL/pgSQL)

Référence : BARO Johanna, 2015, « Données carroyées à 200 mètres sur le bâti en France », <http://hdl.handle.net/TEST/10025>, Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux [Distributeur] V1 [Version]

3. <http://professionnels.ign.fr/bdcarthage>

2.1.2.2 Evaluation du croisement des variables de population et de bâti

Si la question de la sélection des indicateurs les plus pertinents est repoussée au moment de leur exploitation, nous proposons ici de donner des éléments sur la pertinence de leur croisement avec les données de population. L'idée étant de travailler à partir de ces deux sources de données différentes, il semble important de s'assurer au préalable de la qualité de leur croisement. Nous considérons ici la population avec la variable S_{morc} pour évaluer le croisement entre les deux thématiques. Et pour évaluer la pertinence du croisement, nous nous tournons vers les cas extrêmes de valeurs très élevées ou nulles de population et de bâtiments. Il est en effet très difficile d'évaluer tous les cas intermédiaires puisque nous ne disposons pas d'informations sur les taux d'occupation des bâtiments.

La situation extrême la plus fréquemment observée est celle des mailles qui possèdent *une population et une surface bâtie nulles*. Cette situation correspond à toutes les surfaces non bâties qui sont associées à des types d'occupations des sols que nous n'étudions pas ici. Le très grand nombre de ces mailles, s'il n'est pas problématique en termes de qualité du croisement, peut l'être en termes de traitement des données. Nous verrons dans la suite qu'une solution simple consiste à écarter systématiquement des traitements les mailles associées à une surface bâtie nulle.

A l'opposé, la situation qui n'est pas observée est celle des mailles possédant *une surface bâtie nulle et une population non nulle*. Cette situation serait contradictoire. Le croisement des deux thématiques de données issues de sources différentes est cohérent pour cette condition : toute la population résidente est associée à un lieu potentiel d'habitation.

Parmi les mailles d'intérêt, on retrouve des mailles associées à *une population nulle et une surface bâtie non nulle*. Cette situation est attendue lorsque les bâtiments sont de grandes emprises industrielles associées à de grandes surfaces bâties. Elle est plus problématique lorsque le croisement concerne de petits bâtiments. Ce cas de figure peut correspondre à des zones de résidences secondaires, mais on ne dispose pas de variables suffisantes, comme le taux d'occupation des logements, pour pouvoir le confirmer.

Enfin, parmi les mailles d'intérêt on retrouve des mailles associant *des valeurs très élevées de population à des valeurs intermédiaires de surface bâtie*. En recoupant ces caractéristiques avec des informations externes issues de la BD Topo® non transformée ou de Google Maps, on observe que ces mailles correspondent à des espaces occupés par des projets immobiliers associés à de grandes parcelles cadastrales, par exemple des grands ensembles. La figure 2.5 illustre cette situation pour le quartier de Parilly à Bron, sachant que les valeurs exactes de population sont précisées pour les mailles contenant des bâtiments identifiés comme des grands ensembles. Ces valeurs sont très élevées lorsqu'une maille inclut entièrement ou presque un bâtiment ce qui se justifie

en partie par la hauteur des immeubles. Des effets d'agrégation semblent toutefois apparaître lorsque les immeubles sont à cheval sur plusieurs mailles. La population est majoritairement affectée à une maille, mais pas nécessairement celle contenant majoritairement le bâtiment à cause du processus de localisation des ménages à partir du cadastre. Cette situation produit un effet bruité ou mité sur la grille de population qui s'avère dans la suite des travaux problématique pour identifier ce tissu urbain particulier.

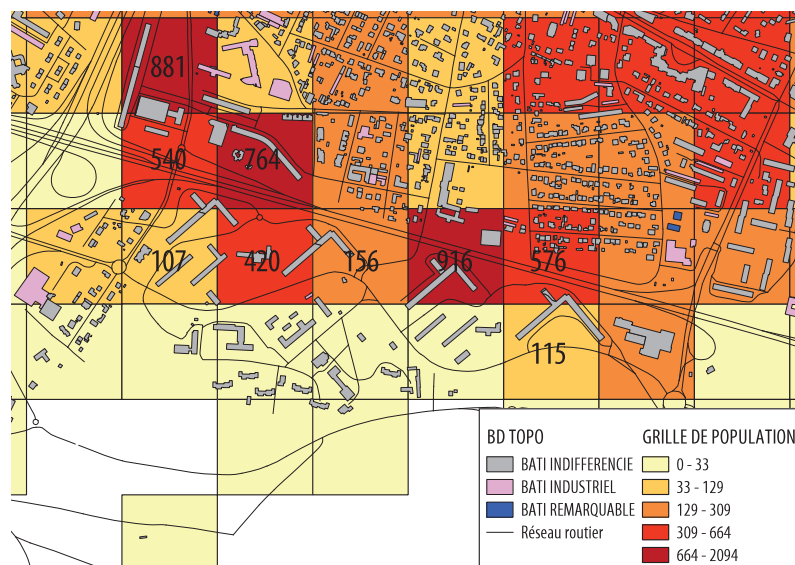


FIGURE 2.5 – Répartition de la population sur la grille de population à maille de 200 mètres de côté pour le quartier de Parilly (Bron, 69)

Le croisement de ces deux indicateurs ne montre donc pas de contradiction empêchant d'utiliser conjointement ces données issues de sources différentes. En complément de ces points relatifs au croisement des variables, nous donnons dans l'encadré 2.4 quelques précisions de qualité sur le calcul des indicateurs à partir des bâtiments de la BD Topo. Les différents points évoqués ici indiquent que les problèmes de modélisation géométrique ou de croisement ne concernent qu'un nombre marginal de mailles qui ne remet pas en cause la pertinence des indicateurs. Il est simplement important de garder ces points à l'esprit lors des analyses menées pour chercher à quantifier les tissus urbains.

2.1.3 Discussion

Prise en tant que thématique d'étude complémentaire à la densité de population, la densité de surface bâtie permet de décrire les espaces urbains selon leur composition en tissus urbains. Ces deux thèmes offrent le moyen de décrire la morphologie au delà de la simple caractérisation des formes de contenant et de contenu. Ils permettent

de caractériser des fonctions (résidentielle/activité) données par la présence/absence de la population, mais aussi des niveaux de hiérarchie et de centralité donnés par les variations de densités. On peut toutefois imaginer que de nombreuses thématiques de données supplémentaires permettraient de caractériser les tissus urbains et les organisations urbaines de façon bien plus complète.

Encadré 2.4 (*Qualité du calcul des indicateurs sur le bâti*)

Sélection des bâtiments Comme évoqué dans la section précédente, les calculs de surface bâtie n'ont pas été effectués sur l'ensemble des bâtiments de la BD Topo®. Cependant la couche « bâti indifférencié » est un peu problématique car elle contient des emprises qu'on aimerait pouvoir filtrer pour une bonne caractérisation des tissus urbains. En effet, l'intégration de grandes emprises de type parking peut faire augmenter considérablement la surface bâtie par maille. Dans des cas extrêmes, mais heureusement rares, ces constructions peuvent faire monter la surface bâtie par maille au maximum possible, à savoir 40000 m². Cette situation est problématique dans la mesure où nous ne nous intéressons pas à la surface totale artificialisée (auquel cas il faudrait en plus prendre en compte l'emprise des voies de communication). C'est la surface des emprises bâties présentant des fonctions résidentielles ou des activités économiques qui nous intéresse pour caractériser des types de tissus urbains et notamment des types d'habitat.

Effets d'agrégation La modélisation géométrique des bâtiments de la BD Topo® peut poser des problèmes de qualité pour les indicateurs de diversité des formes bâties. Lorsque les bâtiments sont agrégés à l'échelle de l'îlot dans des zones hyperdenses, ces indicateurs ont un sens différent. On aimerait pouvoir observer des quartiers à forte densité de surface bâtie et à forte régularité lorsque les surfaces des constructions sont identiques. Or pour ces quartiers qui peuvent par exemple correspondre à de l'habitat en maisons mitoyennes, la régularité est perdue. Au lieu d'observer de nombreuses petites formes de même taille, on observe un ou quelques grands blocs à l'échelle des îlots qui ne sont pas forcément réguliers et dont l'organisation sera entièrement déterminée par la forme du réseau viaire.

Effets de désagrégation La désagrégation sur les grilles régulières des bâtiments lorsqu'ils intersectent plusieurs mailles peut aussi avoir des effets potentiellement gênants lors de l'utilisation des divers indicateurs. La méthode utilisée pour agréger les bâtiments et calculer leur surface fait qu'une maille de 200 mètres de côté peut contenir une valeur de surface bâtie ne correspondant même pas à un bâtiment entier. Ces très faibles valeurs de surface pourront être problématiques pour une étude statistique. Il pourrait être préférable de lisser ces valeurs en fixant automatiquement à zéro les mailles de surfaces bâties inférieures à la taille moyenne d'une habitation individuelle (de 111 m² selon l'enquête logement de 2006 de l'INSEE).

Des mesures de densité encore plus variées, et notamment relatives à l'emploi ou à la spatialisation des activités, permettraient de définir une grille de lecture de la ville multifonctionnelle telle que décrite dans (Batty et al., 2004). La mixité et la diversité des espaces urbains sont des composantes d'organisation qui renvoient principalement aux fonctions et aux activités. Mais comme souligné par certaines études (Stead et Marshall, 2001; Ewing et Cervero, 2010), ces composantes sont corrélées à la densité, qui peut jouer comme un intermédiaire exprimant implicitement des caractéristiques de mixité et de diversité. De plus, ces caractéristiques relatives aux fonctions peuvent être approchées par des caractéristiques morphologiques permettant de déduire de l'information sur la variation d'utilisation des sols. Des mesure de fréquence, d'entropie ou de dimension fractale (Srinivasan, 2002; Sahraoui et al., 2014) calculés sur les données concernant le cadre morphologique peuvent alors être utilisés pour fournir une image approchée de cette diversité.

D'autres mesures relatives à l'utilisation et aux performances des réseaux de transport permettraient de définir une grille de lecture affinée du fonctionnement des systèmes de transport (Stead et Marshall, 2001). Observé par exemple à travers des flux ou des pratiques de mobilité (Schwanen et al., 2004; Le Néchet, 2011), ou à travers des mesures d'accessibilité (Cervero et Kockelman, 1997; Ewing et Cervero, 2010), ces mesures offrent une entrée relationnelle qui permettrait de renseigner sur les structures d'interactions entre les lieux. Complémentaire à la structure morphologique de la ville, cette entrée est sans doute la plus difficile à caractériser du point de vue de la disponibilité des données, bien que de nouvelles sources, issues par exemple des traces numériques (télécommunication, billétique) des individus (Louail et al., 2014) soient en train d'émerger.

La disponibilité de données à l'échelle de la France, sur un maillage compatible avec les grilles régulières à maille de 200 mètres de côté, constitue une contrainte forte pour mener une caractérisation aussi détaillée. Naturellement les autres indicateurs socio-économiques diffusés par l'INSEE auraient pu constituer une source d'information pertinente pour enrichir la caractérisation des tissus urbains. Le calendrier de diffusion quelque peu chaotique de ces données (cf. encadré 2.1), nous a cependant empêché de les intégrer à ce travail de thèse. Nous n'excluons cependant pas de les considérer dans une perspective d'enrichissement à moyen terme.

La représentation qui peut être construite à partir des données de densité est donc une représentation morphologique qui, nous l'avons vu, trouvera son intérêt pour des études où la densité joue un rôle déterminant en tant que conditionnant des fonctions et des interactions entre lieux. La morphologie et le support de description des espaces urbains que nous proposons constitue un cadre de lecture simple et unidimensionnel, et qui présente l'intérêt de pouvoir être facilement croisée et enrichie par les autres thématiques de données citées.

2.2 Le choix d'une nomenclature de tissus urbains

Nous présentons dans cette section des travaux de classification préparatoires pour déterminer une nomenclature des tissus urbains à partir des données de population et de bâti introduites dans la section précédente. A ce stade, l'objectif est avant tout de déterminer les catégories de tissus urbains potentiellement observables à partir de telles données de densité, tout en étant conformes aux caractéristiques d'organisation du modèle morphologique présenté en fin de chapitre 1.

Indépendamment du choix des variables d'intérêt, la question de la quantification des espaces urbains est liée au référentiel spatial des variables. Le maillage contient une structure morphologique, laquelle doit être adaptée au phénomène étudié. Pour faciliter l'identification de ces tissus urbains, nous proposons d'adapter dans un premier temps le maillage des grilles à mailles carrées de 200 mètres de côté, et de la transformer en un maillage radioconcentrique. Nous reviendrons aux grilles régulières lors du chapitre suivant, mais nous choisissons de le transformer dans un premier temps afin de faciliter l'analyse des données de densité. Nous choisissons donc pour ces travaux le maillage spatial le plus pertinent pour identifier les catégories de tissus urbains, en reconnaissant l'existence de modèles d'organisation spatiale qui influenceront la répartition des densités mesurées.

Ce maillage permettra d'identifier plus facilement des seuils pour réaliser des travaux de classification à partir de la construction manuelle d'un arbre de décision. L'objectif n'étant pas encore à ce stade de présenter un traitement exhaustif des territoires urbains français, nous chercherons à classer les tissus urbains de trois agglomérations test : Besançon, Tours et Nancy.

2.2.1 Modéliser l'organisation urbaine par une structure radioconcentrique

Nous proposons ici de construire des classifications des tissus urbains en imposant aux données une première structuration qui intègre des connaissances extérieures de nature thématique et de nature morphologique sur le rapport entre urbanisme et densité. Ces connaissances *a priori* représentent les caractéristiques générales d'organisation déterminantes pour qualifier les structures et construire le modèle morphologique proposé à la fin du chapitre précédent.

Nous choisissons d'intégrer un *a priori* thématique qui opère par le choix des variables de description et de classification des mailles en fonction des objets qu'elles regroupent. L'idée est alors de reprendre les deux thèmes de densité de contenant et de contenu afin de confronter la densité de la population résidente à celle du bâti, et d'évaluer la texture du tissu bâti en fonction de la forme et de la répartition spatiale des

bâtiments. Trois indicateurs spatialisés sur des mailles régulières ont donc été retenus :

- la population résidente (P) ;
- la somme des surfaces bâties (S_{morc}) ;
- la surface moyenne des bâtiments (S_{moy}).

L'*a priori* morphologique est contenu, lui, dans la géométrie du référentiel spatial qui sert d'accroche aux données et qui les met en forme pour cette recherche thématique via des profils orientés de densité. Nous détaillerons dans ce qui suit la démarche exploratoire qui vise à fixer les objectifs de traitement, en neutralisant dans un premier temps la structure spatiale qui sera déterminée par le maillage radioconcentrique.

2.2.1.1 *A priori* morphologique : la structure radioconcentrique

De nombreuses études attachées à décrire la structure des villes se font en utilisant des représentations par profil de densité. Des cercles concentriques sont généralement construits autour d'un centre, qu'il soit historique et patrimonial – car la patrimonialisation conditionne la préservation des bâtiments anciens –, économique – le CBD – (Bertaud, 2004; Guérois et Pumain, 2008; Longley et Mesev, 2002) ou démographique – le maximum de densité de population (Parr, 1985). En général, un seul profil moyen résume l'organisation et la forme urbaine de l'agglomération dans un rapport d'opposition centre-périphérie.

Plutôt que de représenter les agglomérations par ce profil qui masque le caractère anisotrope de l'espace, nous choisissons de réaliser une série de profils sur des secteurs de couronne autour du centre historique des agglomérations étudiées. Nous pouvons ainsi observer la structure urbaine en la confrontant à un modèle sectoriel (Hoyt, 1939), tout en conservant le rapport d'opposition centre-périphérie qui se traduit par un modèle d'organisation par gradients décroissants (Clark, 1951). Les profils mettent en lumière les ruptures induites par l'histoire, les infrastructures ou les contraintes topographiques, et donc les dissymétries structurelles de l'agglomération. De plus, ils permettent de traiter les données en tenant compte des relations de voisinage afin de mieux repérer des discontinuités de faible amplitude, entre des zones proches pour certains indicateurs.

Le choix d'une grille radioconcentrique s'adapte à la fois à notre cadre théorique et applicatif, et à une contrainte statistique. En effet, les profils construits sur une telle grille (cf. encadré 2.5) fournissent une information sur la structure de l'occupation du sol facile à lire dans nos cas d'applications qui correspondent à des modèles de ville européenne (Louail, 2010). Ils révèlent facilement les points de rupture ou seuils pour les indicateurs locaux de densité et de surface moyenne. Ces seuils traduisent empiriquement l'*a priori* mis sur la recherche des différenciations internes produites lors du développement des espaces urbains : les formes à identifier catégorisent le tissu urbain

selon des modes d'occupation du sol produits par la densification et l'extension de la ville, et repérés par rapport au centre historique. Sur ces profils de densité décroissante, la limite de l'agglomération apparaît comme le premier long segment de valeur zéro ; les variations qui la précèdent informent sur les transitions entre les différents tissus urbains composant l'agglomération.

En outre, sur les grilles régulières ayant une maille de 200 mètres de côté, les agglomérations et leurs structures peuvent paraître assez bruitées. A une telle résolution, les espaces intra-urbains non bâtis introduisent une variabilité locale qui donne cette texture bruitée. En construisant les profils sur une grille radioconcentrique, nous effectuons un lissage à pas variable en fonction de la distance au centre. Les mailles périphériques couvrent des portions de territoire plus grandes que les mailles centrales, rendant ainsi plus robustes les mesures de densité de population et de bâti dans les zones peu denses. La variabilité locale est conservée à proximité du centre afin de pouvoir saisir la structure en zones denses. Mais elle est lissée en s'éloignant du centre, où une représentation simple et homogène nous suffit pour caractériser les zones à tendance rurale.

Encadré 2.5 (*Construction des grilles radioconcentriques*)

Les grilles radioconcentriques sont centrées conventionnellement sur la mairie de la commune-centre des agglomérations. Elles contiennent une série d'anneaux concentriques espacés de 200 mètres et coupés par 71 lignes radiantes, qui couvrent ainsi de façon détaillée les agglomérations sur l'ensemble de la zone densément urbanisée jusqu'à la campagne périphérique. Les statistiques relatives à la population et à la surface bâtie issues des grilles régulières sont reprojetées sur ces secteurs par une intersection spatiale des mailles régulières et irrégulières, pondérée par le ratio des surfaces des mailles croisées. Elles sont exprimées en densité, et données par rapport à la surface variable des mailles.

Sur cette nouvelle grille, la construction des profils multivariés, qui distribuent la population, la surface bâtie et la surface moyenne par rapport à la distance au centre est immédiate : chaque profil correspond au vecteur de valeurs donné par la suite des secteurs présents sur une radiale de la grille.

2.2.1.2 Hypothèses sur l'organisation urbaine

Nous posons l'hypothèse que le couplage des trois variables (population, surface bâtie et surface moyenne des bâtiments) permet de réaliser une partition totale de l'espace en quelques grands types de tissus urbains identifiables à des catégories classiques de paysages urbains. Ces catégories de paysages urbains se révèlent en appliquant une grille d'analyse sur les densités de population et de surface bâtie et sur leurs positions observées à l'aide des profils. Nous considérons alors une nomenclature en six classes qui répond aux différents rapports d'opposition introduits dans le tableau 1.1 pour caractériser l'organisation des tissus urbains en termes de hiérarchie, de fonction, de

forme et de relation :

- **Habitat individuel dispersé** : Ce tissu urbain est caractérisé par une dominance des espaces vides et une position périphérique. Il est composé majoritairement de maisons individuelles repérables par la taille moyenne de leur emprise au sol, mais peut aussi contenir des bâtiments de plus grande taille comme des fermes. Il se distingue du tissu urbain suivant par ses plus faibles densités de population et de surface bâtie, auxquelles il est possible d'associer une organisation spatiale marquée par la dispersion des formes bâties.
- **Habitat individuel groupé** : Cette catégorie de tissu urbain proche de la précédente est définie dans la continuité de celle-ci. Les formes de bâtiments correspondent aussi majoritairement à des maisons individuelles, mais qui sont organisées selon des configurations plus denses en population et en surfaces bâties. Associé à une position moins périphérique, ce tissu urbain regroupe des zones d'habitat groupé observables dans des villes et villages, des zones aux organisations régulières (lotissement, linéaire routier) ou irrégulières.
- **Habitat mixte et dense** : Dans cette catégorie, toujours définie dans la continuité de la précédente, la proportion de vide continue de diminuer. Les zones d'habitat sont plus denses en population et en surface bâtie, et localisées à des positions plus centrales. Ce tissu urbain est caractérisé par un mixte de constructions individuelles et collectives qui donnent une surface moyenne intermédiaire. Il regroupe des formes d'habitat de type pavillon, maison de ville, immeuble collectif bas ou intermédiaire. Si ce tissu est dominé par des fonctions résidentielles, les formes de bâti et la position plus centrale laissent à penser que ce tissu peut potentiellement accueillir un nombre plus important d'activités économiques.
- **Habitat collectif** : Cette catégorie en rupture avec les précédentes est caractérisée par la dominance de formes bâties de type immeuble collectif. Associée à une plus grande proportion d'espaces vides que la précédente catégorie, elle est pourtant marquée par une très haute densité de population qui révèle la présence d'un développement vertical. La position de ce tissu urbain est un peu plus variable que pour les autres catégories, mais elle reste toutefois relativement centrale, ou plus centrale que celle des tissus d'habitat individuel groupé et dispersé. Ce tissu urbain pourrait être qualifié autrement comme la catégorie des grands ensembles.
- **Centre historique** : La principale caractéristique de cette catégorie de tissu est l'aspect résiduel des espaces vides. L'espace est structuré par la présence d'immeubles à forte emprise au sol qui permettent d'accueillir une forte densité de population, quoique moins forte que celle de l'habitat collectif. Ce tissu urbain peut être associé aux zones d'urbanisation historiques et centrales dans lesquelles on peut supposer la présence d'une importante mixité en termes de fonction résidentielle et d'activités économiques.

- **Zones d'activités** : Cette catégorie se distingue de toutes les autres puisqu'elle regroupe toutes les zones bâties à usage majoritairement non résidentiel. Elle se caractérise principalement comme étant un tissu urbain à densité de surface bâtie intermédiaire à importante, mais à densité de population nulle ou presque. Cette catégorie complémentaire ne peut pas être associée à une position spécifique dans l'organisation urbaine. Elle regroupe en fait des emprises bâties qui peuvent être très variables en termes de fonctions, qui peuvent être liées à l'administration et aux services publics, aux activités commerciales ou industrielles.

Les hypothèses de détection de ces catégories de tissus à partir des 3 variables sur les grilles radioconcentriques sont résumées dans le tableau 2.2. Ces hypothèses sont déterminées en fonction des villes de notre échantillon d'étude qui présentent la caractéristique commune d'être des villes de taille moyenne (De Roo, 2005). Il est à noter que la représentation vectorielle des bâtiments dans la BD Topo® procède à des agrégations dans certaines zones densément construites comme les centres historiques. Ceci influence la mesure de l'indicateur de surface moyenne : les bâtiments du centre historique présentent généralement de grandes surfaces moyennes dans la BD Topo®.

TABLE 2.2 – Hypothèses sur les valeurs relatives des indicateurs qui permettront de détecter différents tissus urbains

	Position	Population	Surface Bâtie	Surface moyenne
Habitat individuel dispersé	Périphérique++	–	–	–
Habitat individuel groupé	Périphérique+	–	–	–
Habitat mixte et dense	Centre-	+	++	–
Habitat collectif	Centre-	+++	+	+
Centre historique	Centre+	++	+++	++
Zones d'activités	Variable	–	+++	++

Le tableau 2.2 synthétise ainsi une connaissance qualitative des formes urbaines, que nous traduirons en valeurs quantitatives dans la suite de ce travail. Ces formes occupent généralement une position-type que nous ne cherchons pas à définir mais qui permet de penser que deux seuils de densité relativement proches en valeur mais associés à des positions bien distinctes différencient l'espace urbain de façon significative.

2.2.2 Classification des tissus urbains par règles de décision

Travailler à partir de grilles radioconcentriques permet de réorganiser l'information statistique pour construire les classifications en tissus urbains selon une méthodologie simple et robuste par règles de décision. Pour classer les mailles des grilles radioconcentriques selon des types de tissus urbains, nous avons déterminé à partir des profils plusieurs seuils pour chaque variable, en utilisant la connaissance de la trame bâtie des agglomérations. L'objectif est ici de démontrer qu'en se basant simplement sur la

structuration radioconcentrique des données, et *a minima* sur la connaissance de cette trame bâtie, il est possible d'identifier des seuils fournissant des catégories de tissus urbains conformes aux hypothèses émises dans le tableau 2.2 et pertinentes d'un point de vue statistique.

2.2.2.1 Les règles de décision

Une fois combinés sous forme d'arbre de décision (figure 2.6), ces seuils nous permettent de tester l'appartenance de chaque maille irrégulière à l'une des catégories du tissu urbain. Les seuils, sans être universels, présentent une certaine robustesse, puisqu'ils sont en lien avec des aspects physiques des villes européennes considérées : taille d'un bâtiment, d'un ensemble de bâtiments, surface moyenne par habitant, etc. Ils ne sont pas appris statistiquement, car cette approche nécessiterait d'effectuer un apprentissage supervisé, et poserait ainsi la question de la constitution d'une base d'apprentissage, à ce stade aussi problématique que la définition des seuils par catégories. Leur détermination à partir de l'analyse des profils permet de prendre en compte des spécificités régionales, telles que les formes architecturales dominantes, ou la nature de la trame viaire.

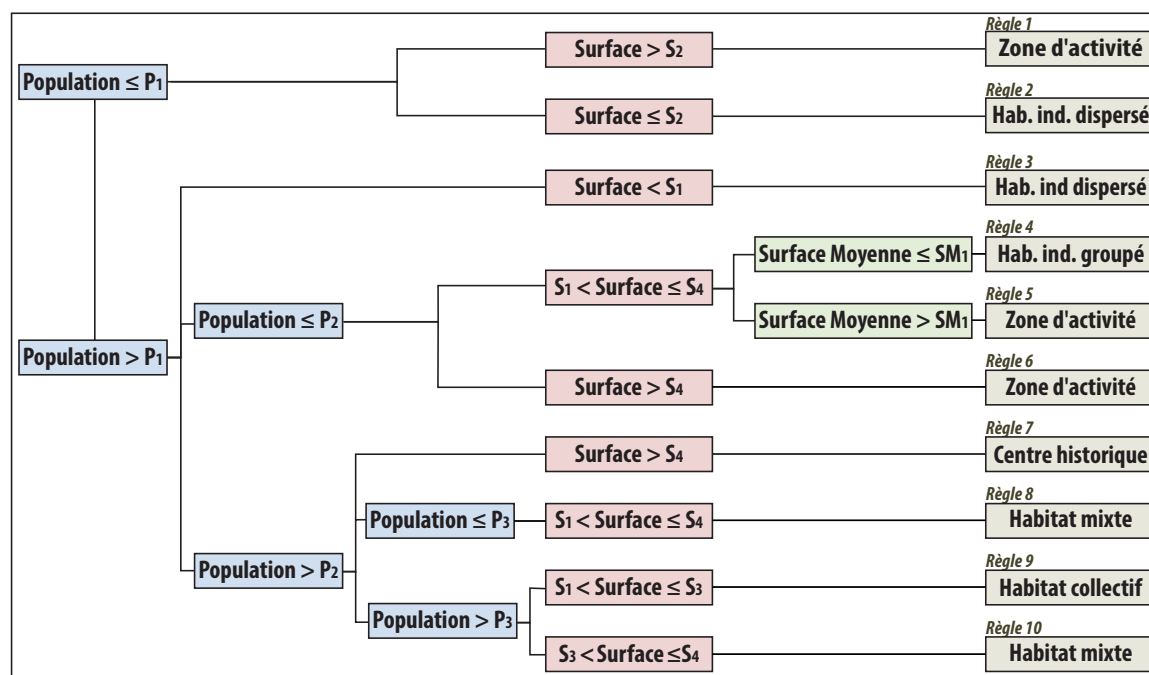


FIGURE 2.6 – Règles de décision pour la classification des tissus urbains

Les seuils ne concernent que les indicateurs de densité de population, surface bâtie et surface moyenne des bâtiments, puisque la transformation des grilles régulières en grilles radioconcentriques intègre un lissage en fonction de la distance au centre. Ainsi, mettre les données sous forme de profils caractérisés par des gradients globalement décroissants

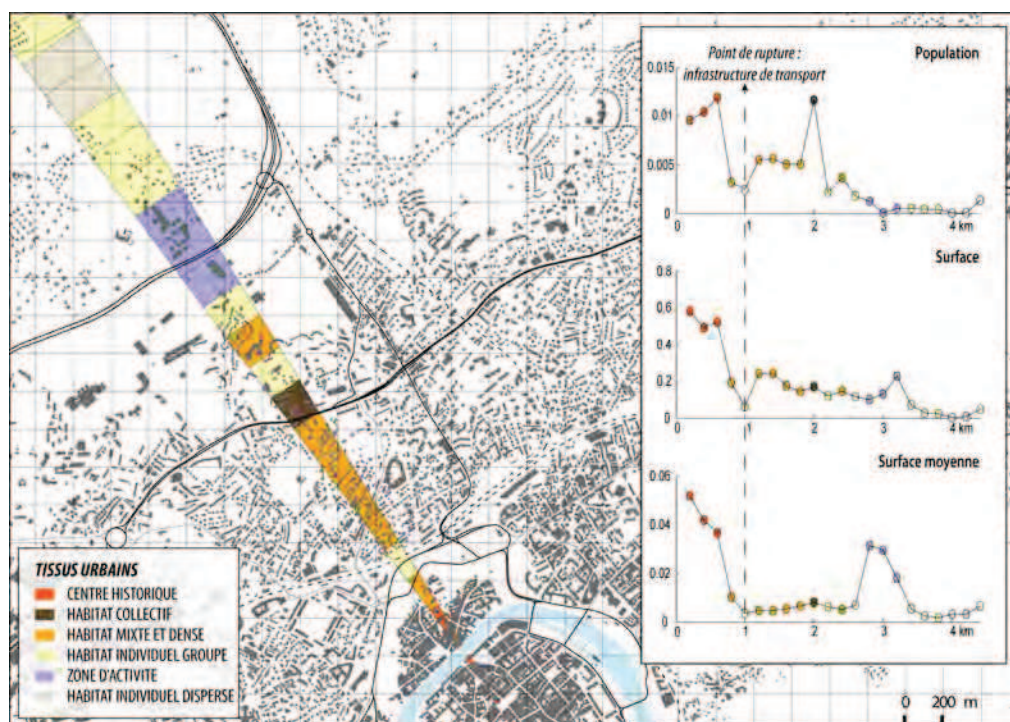
permet de repérer plus facilement les transitions et donc les valeurs seuils pour les indicateurs déterminant les types de tissus urbains à identifier. Selon les hypothèses émises dans le tableau 2.2, ces seuils correspondent à des points de rupture sur les profils de chaque variable en supposant que les transitions apparaissent de façon assez marquées entre chaque catégorie. Dans la pratique, ces transitions sont toujours assez marquées pour au moins une des trois variables utilisées.

Pour illustrer ce propos, la figure 2.7 présente deux exemples de profils dans deux directions différentes sur l'agglomération de Besançon. Leur coloration correspond aux types de tissus urbains qu'il est possible d'identifier selon les points de ruptures et donc les règles de décision. Nous pouvons observer des variations correspondant à des points de rupture marqués, comme la coupure induite par la présence du Doubs (figure 2.7b) ou par des infrastructures de transport (figure 2.7a). Ces profils présentent aussi des gradients moins marqués, associés à des changements progressifs dans l'utilisation des sols comme le passage d'une morphologie d'habitat mixte à une morphologie d'habitat pavillonnaire (figure 2.7b).

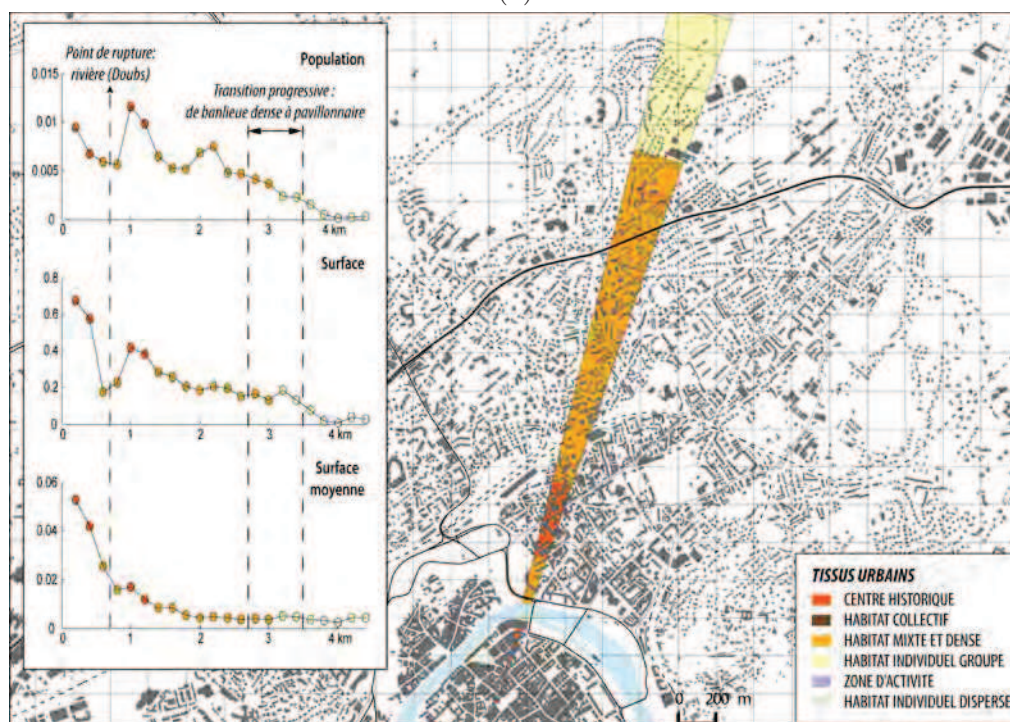
Puisque certaines transitions peuvent être progressives, notamment entre les catégories « habitat individuel » et « habitat mixte », il est parfois nécessaire de faire appel à la connaissance de la trame bâtie de l'agglomération afin de repérer les profils où il sera plus probable d'observer des points de rupture un peu plus marqués. Les valeurs seuils sont alors déterminées en se basant sur les quelques profils les plus pertinents, présentant les ruptures les plus marquées pour discriminer deux catégories de tissus urbains. L'identification de ruptures, ou de zones de changement de concavité sur les profils, présente l'avantage de s'ancrer dans la pratique de l'analyse urbaine. Les profils renvoient à l'organisation des densités urbaines mise en évidence par [Clark \(1951\)](#), et généralement à des types d'urbanisme qu'il est possible de situer dans le temps et dans l'espace : développement des faubourgs, zones pavillonnaires, grands ensembles, etc.

TABLE 2.3 – Seuils pour la classification des tissus urbains selon les agglomérations

		Besançon	Tours	Nancy
Population (hab/maillage de 200m de côté)	P_1	10	10	10
	P_2	108	148	140
	P_3	400	400	400
Surface (m ² /maillage de 200m de côté)	S_1	600	600	600
	S_2	2000	2000	2000
	S_3	8000	10000	10000
	S_4	12000	12000	13400
Surface moyenne (m ² /maillage de 200m de côté)	SM_1	340	440	340



(a)



(b)

FIGURE 2.7 – Profils de densités sur l'agglomération de Besançon : a) profil 47, b) profil 57 (Habillage : BD Topo®, IGN)

La partition de l'espace est construite à l'aide des règles de décision présentées dans la figure 2.6. Pour les trois agglomérations de cette étude, les seuils varient légèrement d'une agglomération à l'autre du fait de leurs tailles différentes et des spécificités locales qui influencent aussi bien les densités de population que la trame bâtie. Le tableau 2.3 résume donc les différents seuils qui ont été utilisés pour identifier les tissus urbains des trois agglomérations. Il a fallu utiliser dix règles de décision pour partitionner totalement les grilles. Cette nécessité d'utiliser plus de règles qu'il n'y a de classes s'explique notamment par la variabilité des objets agrégés dans la classe « zone d'activité » (règle 1 : emprises isolées ou rurales, règles 5 : grandes emprises mêlées à des zones d'habitation, règle 6 : très grandes emprises industrielles). Regrouper les règles donne plus de robustesse aux classes et pallie le manque d'information sur les fonctions des diverses emprises détectées par ces trois règles. De plus, il a fallu intégrer la règle 3 pour filtrer les erreurs de croisement entre les variables relatives à la population et à la surface bâtie, issues de sources différentes.

2.2.2.2 Validation statistique des catégories retenues

Pour vérifier l'homogénéité statistique de la classification produite par la méthode par arbre de décision, et donc de la nomenclature choisie, nous appliquons une méthode purement statistique sur les mailles des grilles radioconcentriques. La CAH, ou classification ascendante hiérarchique (cf. chapitre suivant, section 3.1), permet d'obtenir un ensemble hiérarchisé de partitions, par regroupements successifs des classes qui représentent initialement chaque individu du jeu de données, sans tenir aucun compte des relations géométriques entre les individus. La métrique euclidienne est employée pour mesurer la distance entre individus, tandis que le critère de Ward, qui minimise la variance intra-groupe et maximise la variance inter-groupe, est utilisé pour regrouper les classes.

Pour les trois agglomérations étudiées, les partitions en six classes ne correspondent pas aux six catégories que nous avons détectées par les règles de décision. Il faut descendre plus bas dans l'arbre hiérarchique des partitions pour isoler la catégorie « habitat collectif », confondue avec celle du centre historique à Besançon (figure 2.8) et Nancy, et une partie de la catégorie « habitat mixte » confondue avec l'habitat individuel à Tours. Cette nécessité de descendre plus bas dans l'arbre pour identifier l'ensemble des catégories s'explique par la forme du nuage de points, similaire pour les trois agglomérations : les partitions de l'ensemble compact de points (centré sur l'habitat d'individuel) apparaissent comme les moins stables et les valeurs extrêmes (centre historique, habitat collectif et grandes emprises industrielles), difficilement groupables, se retrouvent isolées dans des classes contenant peu d'éléments.

C'est pourquoi il faut regrouper a posteriori certaines classes se trouvant sur des mêmes branches de l'arbre hiérarchique, afin d'obtenir des partitions où figurent les six catégories de l'approche par arbre de décision. Dans les trois cas, ce sont les catégories

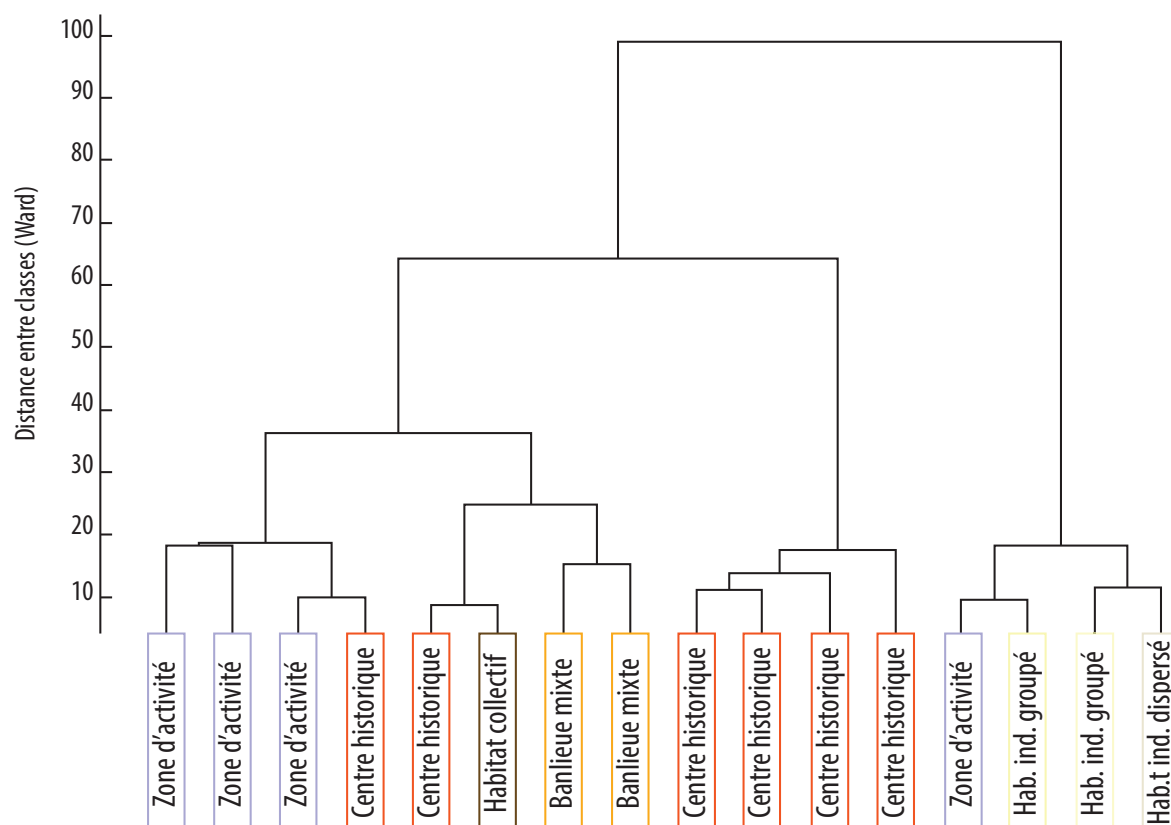


FIGURE 2.8 – Dendrogramme de Besançon (coupé à la partition en 16 groupes)

de zones d'activités et du centre historique qui ont nécessité le plus de regroupements. Leurs densités bâties ou de population affichent une forte variabilité mais les mesures de densité complémentaires ainsi que leur position dans la structure urbaine étant bien déterminées, ces différences peuvent être négligées.

La compatibilité des catégories identifiées par les deux méthodes est mesurée, après regroupement des classes de la CAH, par des matrices de confusion permettant de définir les coefficients Kappa ([Cohen, 1960](#)). Les valeurs de ces coefficients sont 0,85 pour Nancy, 0,83 pour Besançon et 0,82 pour Tours. Elles indiquent une concordance forte entre les deux méthodes de classification pour toutes les agglomérations, et nous assurent de la pertinence et de la bonne différenciation d'un point de vue statistique des catégories retenues.

2.2.2.3 Validation thématique des catégories retenues

L'influence des indicateurs de densité de population et de surface bâtie sur les questions d'aménagement et sur le fonctionnement des systèmes de transport dans l'espace urbain et périurbain doit entraîner la pertinence des classifications construites selon cette nomenclature. Pour vérifier cette pertinence, nous les comparons visuellement aux zonages utilisés pour l'étude des systèmes de transport. Les figures 2.9, 2.10 et

2.11 représentent les secteurs de tirage d'échantillon des enquêtes ménages déplacements (EMD) de Besançon, Tours et Nancy effectuées respectivement en 2004, 2008 et 2012, superposés à nos classifications en tissus urbains.

Les classifications présentées sur ces figures ont été reprojetées sur les grilles régulières pour ne pas perturber la lecture des cartes et pour retourner à l'aspect pratique du carroyage pour la diffusion et le croisement d'informations. La lecture de ces cartes amène à préciser que la classe d'habitat individuel dispersé est en fait à prendre comme une classe complémentaire à l'espace urbain. L'application des règles de décision (figure 2.6) sur un jeu de données construit sans filtrage des mailles vides de bâtiment entraîne le regroupement de ces dernières à cette catégorie de tissu urbain.

Selon la méthode standard CERTU suivie pour réaliser ces enquêtes (Gascon, 2008), les secteurs de tirage doivent être compatibles avec le découpage en IRIS2000 de l'INSEE, contenir 2000 à 5000 habitants et être le plus homogènes possible en termes d'urbanisme et de composition sociale. Ces secteurs sont produits de manière *ad hoc* mais les conditions qu'ils doivent satisfaire sont globalement similaires aux classifications que nous produisons par les règles de décision. Ces secteurs de tirage servent également de base aux zonages utilisés pour l'analyse des pratiques de mobilité, la diffusion des résultats et aussi la modélisation du trafic pour l'évaluation des projets

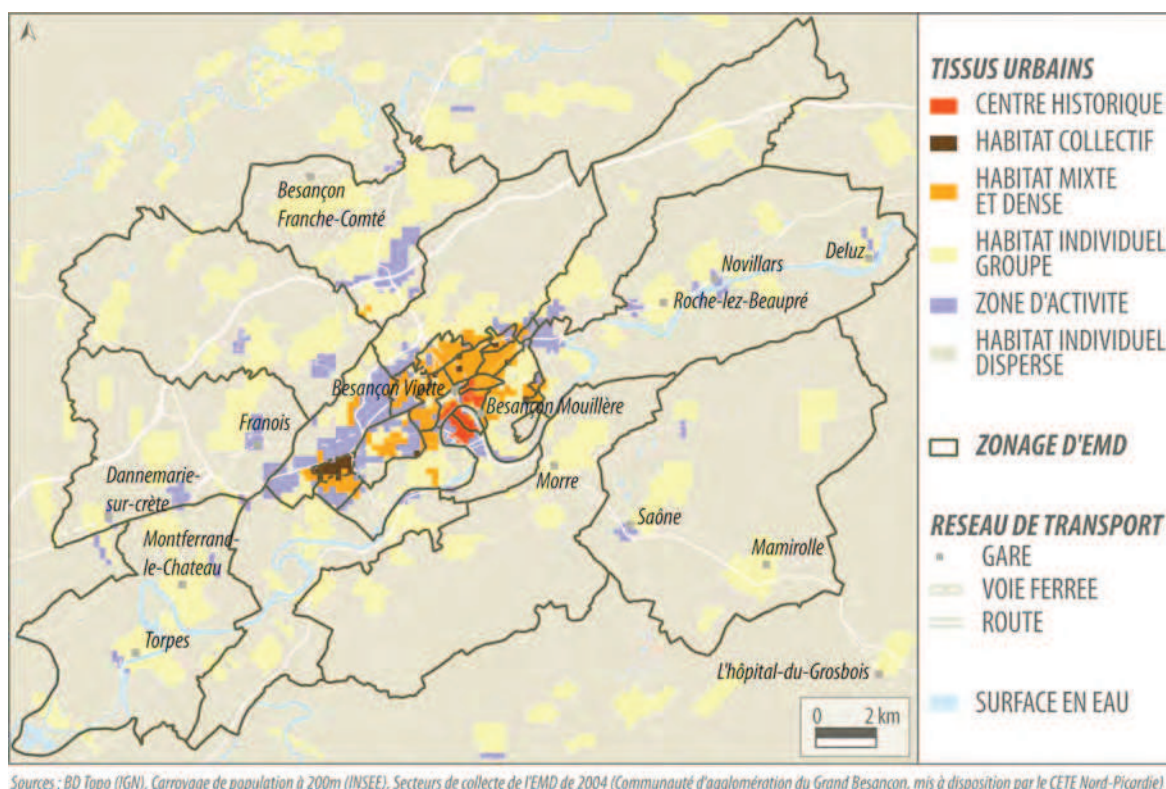
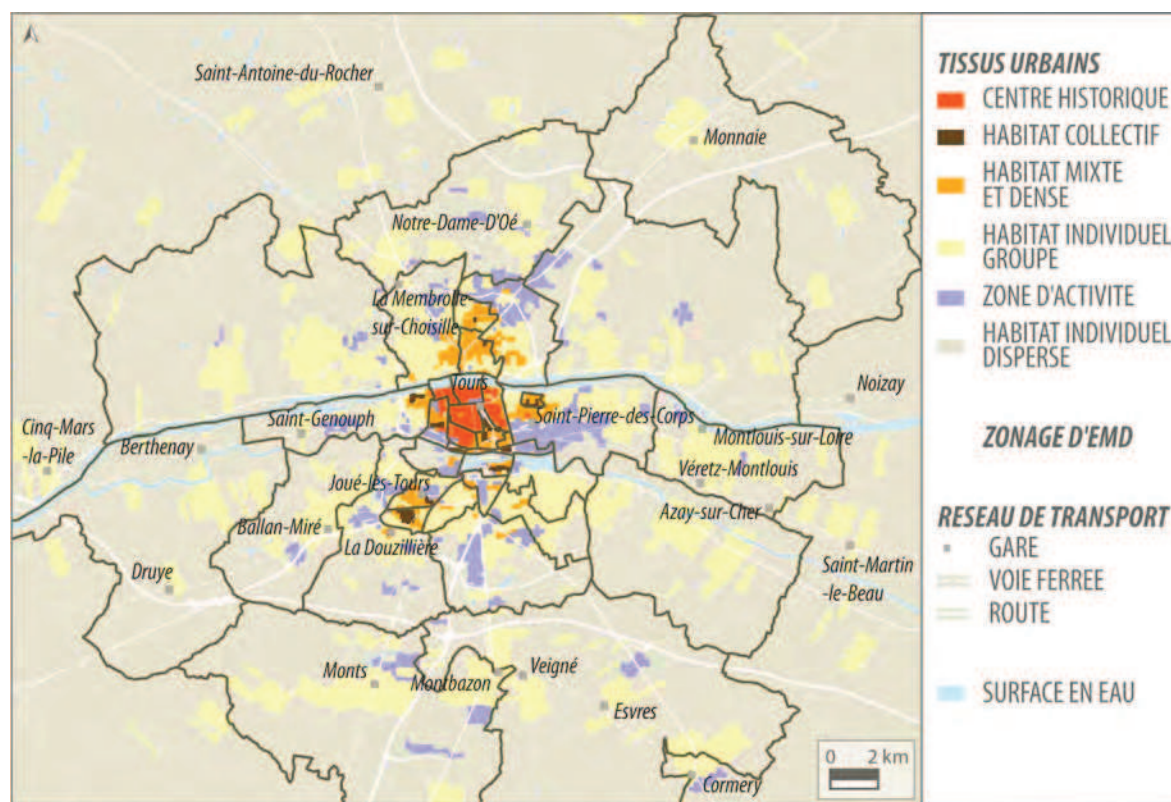


FIGURE 2.9 – Tissus urbains et secteurs d'EMD : Communauté d'agglomération du Grand Besançon

d'aménagement (GACB, 2006; ATU, 2010).

Le découpage des IRIS s'aligne généralement sur les zones densément bâties dans les communes de plus de 10.000 habitants, il est donc normal d'observer une bonne concordance entre les secteurs de tirage et notre classification dans la commune centrale des trois agglomérations, tout particulièrement pour l'agglomération de Besançon. En revanche, les secteurs de tirage des EMD deviennent plus grossiers dans les communes périphériques, non découpées en IRIS. Ils ne différencient pas les zones d'habitat individuel groupé des zones d'habitat dispersé, ni même les pôles secondaires contenant des activités ayant potentiellement un pouvoir attractif en termes de déplacement. Ils regroupent parfois des zones séparées par des infrastructures telles que des autoroutes ou des voies ferrées, qui créent souvent des barrières difficilement franchissables. Les dessertes de telles zones risquent d'être hétérogènes.

Le manque de précision dans les marges des agglomérations empêche de différencier finement les zones résidentielles selon leur densité ou leur situation par rapport au centre, aux nœuds routiers, aux gares, aux zones industrielles et commerciales. Or les zones périurbaines ne sont pas homogènes et différentes pratiques de mobilité peuvent y exister, même si la voiture y est utilisée de façon très dominante. En outre, c'est en périphérie que le bâti et la population évoluent le plus vite, et les comportements de



Sources : BD Topo (IGN), Carroyage de population à 200m (INSEE), Secteurs de collecte de l'EMD de 2008 (SCOT de l'agglomération Tourangelle et sitcat, mis à disposition par le CETE Nord-Picardie)

FIGURE 2.10 – Tissus urbains et secteurs d'EMD : SCOT de l'Agglomération Tourangelle

mobilité peuvent changer à moyen terme, au moins dans certaines zones.

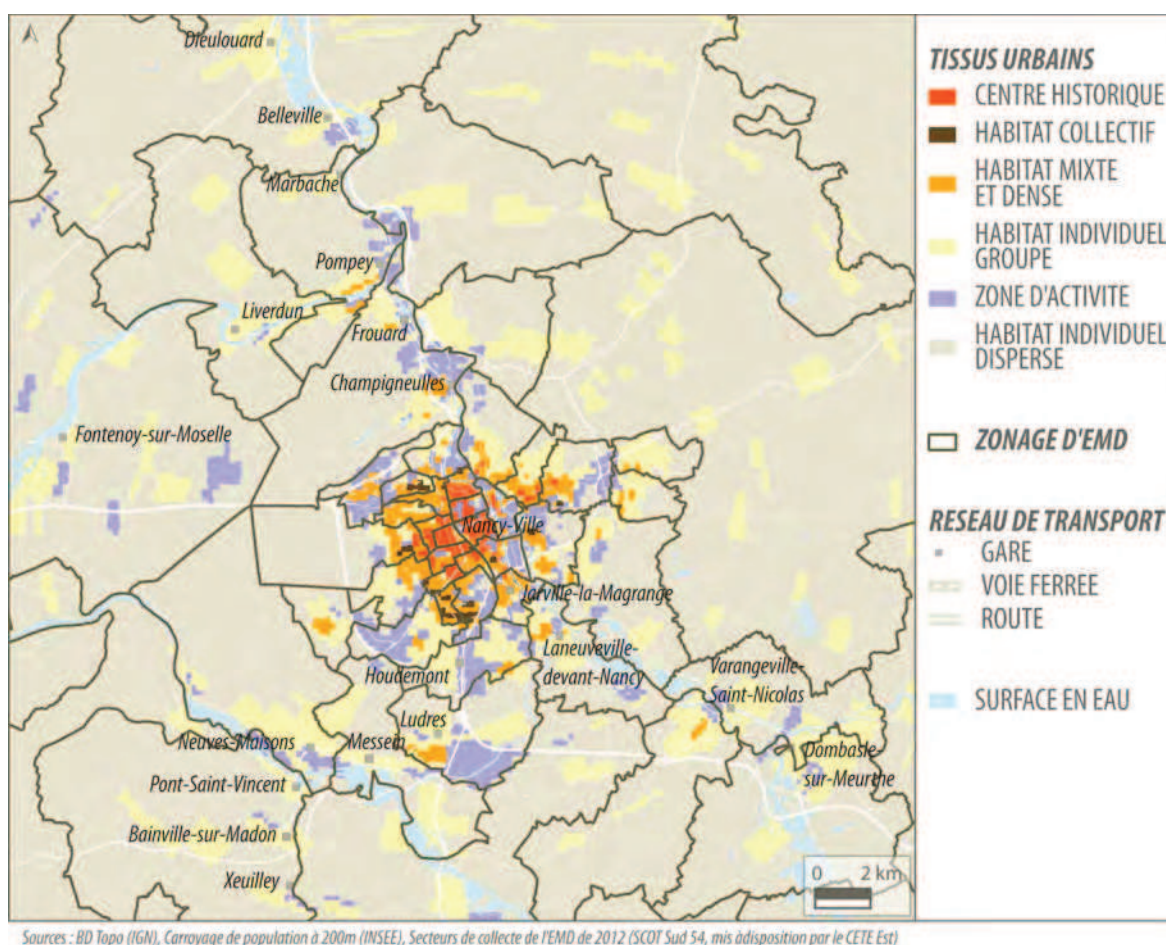


FIGURE 2.11 – Tissus urbains et secteurs d'EMD : SCOT Sud 54

2.2.3 Discussion

L'association sur des données carroyées de la population résidentielle et des données relatives aux surfaces bâties permet de représenter une structure morphologique des agglomérations françaises en termes de densité des bâtiments et des résidents. Les classifications obtenues fournissent une catégorisation de l'espace à l'échelle meso-géographique selon une nomenclature en tissus urbains conforme aux caractéristiques d'organisation présentées au chapitre 1. Ces caractéristiques sont fondées sur les logiques de densification et de spécialisation, de croissance urbaine et de développement centrifuge des réseaux métropolitains. Elles sont donc un support pour comprendre le rôle des réseaux de transport dans le fonctionnement et le développement de l'espace urbain et pour réfléchir aux évolutions possibles en termes de développement urbain et de pratiques de mobilité. Notamment, ces classifications présentent un intérêt pour affiner la connaissance des zones périphériques parfois mal discriminées dans des maillages

de type IRIS ou secteurs de tirage des EMD.

La logique de grille radioconcentrique introduite ici présente l'intérêt d'exposer l'organisation hétérogène des formes urbaines, souvent masquée par les découpages administratifs. Toutefois, imposer une telle grille de lecture des données peut trouver ses limites lorsque l'hypothèse d'organisation morphologique radioconcentrique n'est plus respectée. Un premier recours serait de modifier l'*a priori* morphologique appliqué au maillage géographique, en déformant les grilles régulières selon d'autres contraintes permettant de prendre en compte la structure des réseaux de transport et la présence de centres multiples. Ces déformations permettraient par exemple de traiter le cas des villes plus linéaires, des villes multipolaires ou encore des villes nouvelles.

Un second recours serait de chercher à encoder cet *a priori* autrement dans la structure des données, et plus nécessairement sous la forme d'un maillage géographique irrégulier. Le maillage des grilles régulières peut être conservé sous sa forme originale pour chercher à identifier des tissus urbains. Mais puisqu'il contient un excédent d'information non pertinente (les mailles de surface non bâtie), il faudrait adapter la sélection des données et le modèle statistique employé pour identifier les tissus urbains. En tenant compte des mêmes caractéristiques de densité, nous verrons dans la suite que la modélisation par champ de Markov offre le moyen de transférer cet *a priori* morphologique sous la forme d'un *a priori* probabiliste contenu dans des structures de dépendances spatiales propres à chaque maille et chaque catégorie de tissu urbain.

Ces premiers résultats de classification ouvrent ainsi une réflexion sur l'automatisation du traitement d'un plus grand nombre d'agglomérations. La nécessité de fixer des seuils pour chaque agglomération limite l'opérationnalisation de cette méthode pour un travail exhaustif. Pour autant, les résultats obtenus constituent un premier pas vers cette automatisation, en appliquant un processus de classification supervisée à partir d'une base d'apprentissage qui est maintenant constituée grâce à ce travail.

Conclusion

Ce chapitre nous a permis d'introduire l'approche de quantification sur les grilles régulières que nous proposons pour identifier et étudier des structures urbaines selon des caractéristiques morphologiques. Nous avons fait le choix de considérer une combinaison de mesures de densité de contenant et de contenu adaptées à la lecture morphologique des organisations urbaines. Ce choix découle de l'intérêt que peuvent avoir ces variables pour l'étude de structures urbaines, mais aussi de la disponibilité de ces données à l'échelle de la France. Si la qualification que nous pouvons déduire de ces mesures de densité peut apparaître relativement pauvre, nous avons vu qu'en combinant quelques indicateurs de densité il est toutefois possible de caractériser différents types de tissus urbains pertinents pour comprendre les organisations urbaines

à travers des problématiques de transport et de développement urbain.

Nous disposons d'une connaissance *a priori*, maintenant considérée sous forme de catégories de tissus urbains empiriquement observables à partir des données de densité. Les grilles radioconcentriques utilisées pour déterminer ces catégories constituaient un support intégrant en plus un *a priori* morphologique facilitant l'exploitation des données, ce qui nous a servi à fixer les objectifs de représentation. La suite de ce travail doit toutefois s'articuler autour de méthodes plus automatiques et opérationnelles. Nous verrons dans le chapitre suivant qu'en conservant le maillage des grilles régulières, mais en développant d'autres méthodes de classification (intégrant d'une autre façon l'*a priori* morphologique), il sera possible d'identifier cette même nomenclature de tissus urbains pour un ensemble beaucoup plus important d'espaces urbains.

Classification des tissus urbains par transfert d'apprentissage

Sommaire

Introduction	92
3.1 État de l'art des méthodes de classification	93
3.1.1 Les approches généralistes	93
3.1.2 Les approches par transfert d'apprentissage	101
3.1.3 Les approches spatiales	104
3.2 Transfert d'apprentissage et champ de Markov	115
3.2.1 Modèles de transfert dans le cadre Markovien	115
3.2.2 Estimation de paramètres	117
3.2.3 Expérimentations sur données simulées	122
3.3 Application : la construction d'une base de tissus urbains	126
3.3.1 Mise en œuvre de la base de données	128
3.3.2 Évaluation de la méthode	135
3.3.3 Évaluation de la base de données des tissus urbains	143
Conclusion	154

Introduction

Ce chapitre est consacré à la présentation d'une méthode de classification permettant de formaliser une première représentation des structures urbaines étudiées selon leur composition en tissus urbains. L'enjeu est de présenter une méthode de classification adaptée au traitement exhaustif de divers territoires urbains présents sur le sol français. Il s'agit de mettre au point une méthode qui répond aux critères suivants :

- Les classifications doivent être comparables sur les divers espaces étudiés : il s'agit de reprendre la nomenclature en 6 classes de tissus urbains introduites dans le chapitre précédent, afin de produire des images classées des territoires qui puissent être comparées sans avoir recours à des appariements d'étiquettes.
- La méthode de classification doit être adaptée au traitement de données spatialisées sujettes au phénomène d'autocorrélation spatiale. Nous savons à ce stade que la distribution spatiale des classes n'est pas aléatoire, les valeurs de mailles voisines ne sont pas indépendantes. Il s'agit alors de prendre en compte ces structures spatiales qui impactent les résultats de classification.
- Elle doit tenir compte du coût élevé que représente la construction de jeux d'entraînement labellisés pour lancer l'apprentissage supervisé de chaque espace urbain. Il est donc question d'utiliser au mieux la connaissance disponible sur quelques espaces urbains, et de pouvoir la généraliser à d'autres espaces sur lesquelles nous ne disposons d'aucuns labels.
- Elle doit tenir compte de la variété de distributions statistiques et spatiales rencontrée sur ces territoires. Il est nécessaire de tenir compte des phénomènes d'hétérogénéité spatiale qui peuvent entraîner une dérive des paramètres des distributions statistiques par rapport aux quelques jeux d'entraînement qui seront utilisés.

Après une présentation de l'état de l'art des diverses méthodes de classifications utilisées tout au long de ce travail de thèse, nous introduirons une méthode de classification par transfert d'apprentissage reposant sur une modélisation markovienne. Nous évaluerons dans un premier temps cette méthode de classification sur des données simulées afin de démontrer l'apport du transfert d'apprentissage et de la prise en compte de dépendances spatiales.

Enfin nous détaillerons comment cette méthode a pu être employée sur les données réelles correspondant à certains des indicateurs de densité introduits dans le chapitre précédent pour constituer une base de données des tissus urbains sur 50 aires urbaines. Des expériences illustreront le comportement de cette méthode lors de la constitution de cet ensemble de classifications et montreront son intérêt pour poursuivre ce travail exhaustif sur l'ensemble du territoire français.

3.1 État de l'art des méthodes de classification

Déjà utilisées en fin de chapitre précédent, les méthodes de classification sont au cœur de ce travail de thèse. Automatiser la production des classifications en tissus urbains, produire une cartographie plus détaillée des structures urbaines, ou encore produire une typologie des structures urbaines observées sur le territoire français, tous ces objectifs reposent sur l'utilisation de méthodes de classification. Nous présenterons donc dans cette section un état de l'art des méthodes de classification qui repose sur l'ensemble des méthodes auxquelles nous aurons recours tout au long de ce travail de thèse. Les méthodes présentées sont divisées en trois sections :

- Dans la section des méthodes généralistes nous présenterons des approches probabilistes et non probabilistes qui sont très largement employées dans divers domaines faisant appel à l'apprentissage statistique (*machine learning*), qu'il soit supervisé ou non supervisé. Certaines de ces méthodes, comme la CAH ou les arbres aléatoires seront reprises pour réaliser des tâches annexes de classification au cours de notre travail.
- La seconde section introduit la problématique du transfert d'apprentissage, employée notamment pour améliorer la résolution de problèmes de classification bénéficiant de l'utilisation de connaissances acquises sur un problème similaire. Nous nous intéresserons principalement à la problématique du transfert transductif que nous emploierons dans la suite pour proposer une méthode de classification adaptée à nos objectifs.
- Dans la section des approches spatiales nous présenterons les approches de classification ou de segmentation plus spécifique au traitement d'images ou de données spatialisées. Nous nous focaliserons plus particulièrement sur le formalisme des champs de Markov, employé pour répondre au besoin de prise en compte des dépendances spatiales lors du processus de classification. Nous présenterons les divers algorithmes de classification markovienne auxquels nous aurons recours pour proposer la méthode sur laquelle nous travaillerons.

3.1.1 Les approches généralistes

En apprentissage automatique, une tâche de classification correspond à la recherche d'un ensemble de K groupes d'individus homogènes au sein d'un jeu de données. Ce jeu de données noté $\mathbf{x} = (x_1, x_2, \dots, x_n)$ est composé de p variables continues ou catégorielles (aussi appelées ici covariables), et de n individus. Ces groupes sont eux représentés par une variable catégorielle $\mathbf{y} = (y_1, \dots, y_n)$ (label), qui peut être déterminée selon différentes approches : les approches géométriques non paramétriques qui utilisent des critères de groupement des individus définis pour une métrique donnée ; ou les approches probabi-

listes, qui peuvent être paramétriques ou non paramétriques et qui utilisent des modèles probabilistes basés sur des hypothèses de distribution des données pour prédire \mathbf{y} .

De plus, différents problèmes de classification peuvent être distingués en fonction des conditions d'apprentissage. Nous présenterons ici des méthodes selon la distinction générale entre l'apprentissage supervisé et non supervisé. Les approches de classification non supervisée correspondent à différentes méthodes permettant de retrouver différents groupes codés par la variable de classe \mathbf{y} sans connaissance particulière sur cette variable. Le problème est alors d'apprendre la structure sous-jacente des données et retrouver les valeurs de \mathbf{y} de façon à révéler des groupements « naturels » dans les données. L'apprentissage supervisé nécessite lui un ensemble d'apprentissage $\mathcal{A} = \{(x_i, y_i), i = 1, \dots, n\}$ composé d'exemples labellisés. L'hypothèse de l'apprentissage supervisé est que cet ensemble d'entraînement possède la même distribution statistique que le jeu de données à prédire, ce qui permet de généraliser le classifieur à des données non considérées lors de l'apprentissage.

Dans le cadre de ce chapitre, et de la proposition d'une méthode de classification des tissus urbains, ce sont les approches probabilistes supervisées qui nous intéressent le plus. Mais dans la suite des travaux, nous aurons recours à différentes approches non probabilistes pour définir des modèles d'organisations urbaines à des échelles plus macroscopique. Nous nous intéressons ici à présenter quelques méthodes de référence dans ces deux branches de classification, qui nous serviront à mener à bien ces différentes tâches de classification.

3.1.1.1 Les approches non supervisées

Parmi les nombreuses approches supervisées existantes, nous nous intéressons ici à présenter deux familles de méthodes qui peuvent être déclinées en classifieurs utilisant des critères de modélisation des données et de regroupements variés : les méthodes de classifications hiérarchiques, et les méthodes de classification par modèle de mélange.

Les classifications hiérarchiques

Les méthodes de classifications hiérarchiques constituent une approche non supervisée classique qui ne nécessite pas de connaissance sur les groupes à identifier. Elles se distinguent des approches dites plates telles que l'algorithme k-means (Hastie et al., 2009) qui ont la particularité de produire une unique partition des n individus d'un jeu de données en K groupes ou classes.

En effet, les méthodes hiérarchiques produisent une suite de partitions emboîtées des n individus en $K = 1, \dots, n$ groupes. Deux approches de classification se distinguent alors dans les méthodes hiérarchiques. Les méthodes de classification ascendante hiérarchique (CAH), méthodes sur lesquelles nous nous focaliserons ici, regroupent itérativement les

individus initialement répartis en n groupes. Les méthodes descendantes, elles, divisent itérativement les individus initialement inclus dans un unique groupe. Dans les deux cas les approches fournissent $n-1$ partitions qui peuvent être représentées par un arbre. Cet arbre, aussi appelé dendrogramme (cf. figure 2.8 ou 5.13), permet de visualiser les regroupements d'individus à chaque niveau de la hiérarchie.

Ces méthodes de classification non supervisée reposent simplement sur la définition d'un critère de proximité entre individus (distance euclidienne, de Mahalanobis, etc.) et d'une stratégie d'agrégation ou de division des groupes d'individus (Rencher, 2002). Pour une CAH l'algorithme consiste à regrouper itérativement les deux groupes d'individus les plus proches, selon la minimisation d'un critère d'agrégation D , et ce jusqu'à ce qu'il n'y ait plus qu'un groupe contenant tous les individus. Parmi les critères d'agrégation les plus employés, on retrouve les critères suivants mesurant la distance entre deux groupes A et B :

$$\text{Stratégie du saut minimum : } D(A,B) = \min_{a \in A; b \in B} d(x_a, x_b) \quad (3.1)$$

$$\text{Stratégie du saut maximum : } D(A,B) = \max_{a \in A; b \in B} d(x_a, x_b) \quad (3.2)$$

$$\text{Stratégie du saut moyen : } D(A,B) = \frac{1}{n_a n_b} \sum_{a \in A} \sum_{b \in B} d(x_a, x_b) \quad (3.3)$$

$$\text{Critère de Ward : } D(A,B) = SSE_{AB} - (SSE_A + SSE_B) \quad (3.4)$$

où $SSE_G = \sum_{i \in G} (x_i - \bar{x}_G)^t (x_i - \bar{x}_G)$, G étant un groupe d'individus.

Chacun de ces critères produira un arbre hiérarchique de partitions potentiellement différentes pour un même jeu de données. Tous ces critères possèdent des avantages et inconvénients : le critère du saut minimum aura notamment tendance à être sensible aux erreurs de distances entre individus, mais est adapté pour identifier des groupes aux formes curvilinéaires. De même le critère de Ward possède une certaine robustesse face aux valeurs extrêmes qui seront isolées dans le bas de l'arbre et se regrouperont entre elles plutôt que de rejoindre de larges groupes dans les plus hauts niveaux de l'arbre. (Rencher, 2002)

L'algorithme de la CAH ne permet cependant pas de déterminer le nombre de classes et donc la partition la plus adaptée pour représenter le jeu de données. Ce problème récurrent en classification non supervisée est souvent résolu par des approches heuristiques et graphiques pour les algorithmes non probabilistes comme la CAH ou l'algorithme k-means. Le dendrogramme qui représente l'arbre hiérarchique de classification exploite la longueur des branches pour représenter le coût de fusion des groupes. Il fournit par exemple une représentation intuitive pour identifier un niveau de coupe de l'arbre hiérarchique, repéré à l'endroit où les branches sont les plus longues, ce qui indique un coût élevé d'agrégation.

D'autres procédures peuvent être citées (Rencher, 2002). Une première procédure

visé par exemple à déterminer le nombre optimal de groupes en appliquant un seuillage sur les distances $D(A,B)$ (quelle que soit la stratégie) entre l'agrégation de deux groupes. Une seconde, qui est aussi applicable sur des approches non hiérarchiques, vise à identifier le nombre optimal de groupes parmi un ensemble de partitions à partir des distances intra et inter groupes. Enfin d'autres approches peuvent être employées pour valider le nombre de groupes d'une partition, notamment par test d'hypothèse ou par validation croisée.

Modèle de mélange et algorithme EM

A côté de ces méthodes non probabilistes, les modèles de mélanges indépendants constituent une autre approche très largement employée de classification plate non supervisée définie dans un cadre probabiliste. Dans le cadre probabiliste l'idée est d'apprendre la structure sous-jacente des données en posant un ensemble d'hypothèses sur les distributions de \mathbf{x} et de \mathbf{y} .

Notion de variable cachée Il est alors courant de présenter le problème de classification en introduisant deux types de données : les données observées $\mathbf{x} = (x_1, \dots, x_n)$ où $x_i \in \mathbb{R}^p$ et les données non observées $\mathbf{y} = (y_1, \dots, y_n)$, où $y_i \in [0,1]^K$, correspondant aux labels de classes à identifier. La variable \mathbf{y} étant manquante lors du processus d'apprentissage, elle est appelée variable cachée. Ce formalisme est utilisé pour présenter différentes méthodes de classifications. Introduit ici pour les modèles de mélanges, il sera aussi utilisé pour des modèles markoviens.

De façon générale, les modèles de mélanges correspondent à une approche paramétrique dont l'objectif est d'estimer la densité marginale $P(\mathbf{x})$. Cette densité est estimée en supposant que chaque groupe à identifier correspond à une source indépendante et que l'ensemble des données correspond à un mélange de ces K sources. La densité des données observées est donnée par :

$$P(\mathbf{x}) = \sum_{k=1}^K \pi_k f_k(\mathbf{x}; \theta_k) \quad (3.5)$$

où les π_k sont les proportions de mélanges et les f_k sont les composantes du mélange c'est-à-dire les densités de chaque source paramétrées par θ_k . La loi d'un mélange est alors associée à un paramétrage $\psi = (\pi_1, \dots, \pi_K, \theta_1, \dots, \theta_K)$. Il est courant de supposer que le mélange suit une loi multinormale adaptée au traitement de variables continues, bien que d'autres formes de distribution soient envisageables (McLachlan et Peel, 2000). La distribution de chaque classe $k \in \{1, \dots, K\}$ est dans ce cas une gaussienne $\mathcal{N}(\mu_k, \Sigma_k)$:

$$f(\mathbf{x}|\mathbf{y} = k, \theta_k) = \frac{1}{\sqrt{(2\pi)^p |\Sigma_k|}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mu_k)^t \Sigma_k^{-1} (\mathbf{x} - \mu_k) \right) \quad (3.6)$$

où μ_k est le vecteur moyen de dimension p et Σ_k est la matrice de covariance de taille $d \times d$.

Dans le cas d'un modèle de mélange indépendant, on suppose que les classes y_i , $i \in \{1, \dots, n\}$ sont indépendantes et identiquement distribuées (i.d.d.) et que les covariables x_i sont conditionnellement indépendantes aux classes y_i . La loi marginale $P(\mathbf{x}|\psi)$ peut alors s'écrire :

$$P(\mathbf{x}|\psi) = \prod_{i=1}^n \sum_{k=1}^K \pi_k f_k(\mathbf{x}; \theta_k) = \prod_{i=1}^n P(x_i|\psi) \quad (3.7)$$

Le problème de classification non supervisée revient alors à identifier les paramètres du modèle de mélange par la méthode du maximum de vraisemblance. Pour ce problème de données incomplètes, l'algorithme EM (*Expectation Maximisation*) est particulièrement adapté (Dempster et al., 1977). Sa mise en œuvre repose sur la maximisation de l'espérance de la log-vraisemblance complétée $\mathcal{L}(\mathbf{x}, \mathbf{y}, \psi)$ conditionnellement au paramètre courant. Cette log-vraisemblance est donnée par :

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, \psi) = \sum_{i=1}^n \sum_{k=1}^K y_{ik} \log[\pi_k f_k(x_i, \theta_k)] \quad (3.8)$$

L'algorithme EM consiste en partant d'un paramètre initial $\psi^{(0)}$ arbitraire à maximiser itérativement la fonction $\mathcal{Q}(\psi|\psi^{(q)})$:

$$\mathcal{Q}(\psi|\psi^{(q)}) = \mathbb{E}[\mathcal{L}(\mathbf{x}, \mathbf{y}, \psi) | \mathbf{x}, \psi^{(q)}] \quad (3.9)$$

$$= \sum_{i=1}^n \sum_{k=1}^K t_{ik}^{(q)} \log[\pi_k^{(q)} f_k(x_i | \theta_k^{(q)})] \quad (3.10)$$

où $t_{ik}^{(q)} = \mathbb{E}(y_{ik} | x_i, \psi^{(q)})$ est la probabilité *a posteriori* $P(y_{ik} = 1 | x_i, \psi^{(q)})$. Le paramètre $\psi^{(q)}$ est mis à jour à chaque itération selon une procédure en deux étapes :

(E) Calcul des probabilités *a posteriori* :

$$t_{ik}^{(q)} = \frac{\pi_k^{(q)} f_k(x_i | \theta_k^{(q)})}{\sum_{l=1}^K \pi_l^{(q)} f_l(x_i | \theta_l^{(q)})} \quad (3.11)$$

(M) Mise à jour des paramètres (π_k) et $\theta = (\theta_k)$ en maximisant l'espérance de la log-vraisemblance complétée \mathcal{Q} :

$$\pi_k^{(q+1)} = \frac{1}{n} \sum_{i=1}^n t_{ik}^{(q)} \quad (3.12)$$

$$\theta^{(q+1)} = \arg \max_{\theta} \sum_{i=1}^n \sum_{k=1}^K t_{ik}^{(q)} \log f_k(x_i | \theta_k) \quad (3.13)$$

Cette procédure assure que la log-vraisemblance $\mathcal{L}(\mathbf{x}, \psi^{(q)})$ croît à chaque itération et qu'elle converge vers un maximum local de la vraisemblance (McLachlan et Peel, 2000). Un grand nombre de variantes de cet algorithme ont été développées en particulier pour

palier les éventuelles problèmes de convergence vers des maxima locaux de mauvaise qualité, tel que SEM (*Stochastic EM*) (Celeux et al., 1995).

Une fois les paramètres du mélange connus, il est possible de construire une partition en déterminant la classe d'appartenance la plus probable à l'aide d'une règle telle que la règle du *maximum a posteriori* (MAP) :

$$\forall i \in \{1, \dots, n\}, \quad y_i^{MAP} = \arg \max_{k \in \{1, \dots, K\}} P(y_{ik} | x_i, \psi) = \arg \max_{k \in \{1, \dots, K\}} t_{ik} \quad (3.14)$$

Telle que présentée jusqu'ici, la classification par modèle de mélange suppose que le nombre K de classes à identifier est un paramètre connu. Mais comme pour la CAH et toute approche non supervisée, il est courant d'être confronté à la situation où K n'est pas connu et doit aussi être estimé. Dans ce cas, il est possible de faire appel à des critères de sélection de modèle tels que AIC, BIC ou encore ICL¹ (Biernacki et al., 2000).

Si ces modèles sont souvent introduits pour des problèmes de classification non supervisée, l'hypothèse d'indépendance du modèle de mélange fait qu'il est aisé d'étendre le processus de classification à un cadre semi-supervisé (Saul et al., 2006). Dans ce cadre de travail à mi-chemin entre le supervisé et le non supervisé, l'apprentissage d'un modèle statistique repose sur l'utilisation d'un ensemble d'apprentissage composé de m exemples labellisés, et de n exemples non labellisés. Dans le cas des modèles de mélange, la log-vraisemblance peut alors être décomposée en une somme de deux termes : un terme pour les individus labellisés et un terme pour les individus non labellisés. L'algorithme EM reste adapté pour estimer les paramètres du modèle, avec simplement une modification de l'étape E dans laquelle les probabilités *a posteriori* ne sont calculées que pour les individus non labellisés (Côme, 2009).

3.1.1.2 Les approches supervisées

Nous nous intéressons maintenant aux méthodes d'apprentissage supervisé basées sur l'utilisation d'un ensemble d'exemples d'apprentissage $\mathcal{A} = \{(x_i, y_i), i = 1, \dots, u\}$. Nous présentons ici trois méthodes de classifications : la régression logistique, les arbres de décision, et les forêts aléatoires basées sur l'utilisation de multiples arbres de décision.

3.1.1.3 La régression logistique

La régression logistique est un modèle de régression permettant la prédiction d'une variable catégorielle, intégré à la famille des modèles linéaires généralisés (Gill, 2001). Le modèle logistique utilise la fonction *logit* pour modéliser la distribution binomiale

1. Ces trois critères fournissent des mesures de l'ajustement d'un modèle statistique. Ils reposent sur l'utilisation de la vraisemblance du modèle pénalisée en fonction de la complexité du modèle (nombre de paramètres du modèle).

ou multinomiale d'une variable catégorielle. La liaison entre une variable catégorielle \mathbf{y} à K modalités et un ensemble de covariables \mathbf{x} à p variables est donnée par des probabilités conditionnelles exprimant des relations curvilinéaires :

$$\log \left[\frac{P(\mathbf{y} = k | \mathbf{x} = x)}{1 - \sum_{l=1}^{K-1} P(\mathbf{y} = l | \mathbf{x} = x)} \right] = \beta_{k,0} + x\beta_k \quad \forall k \in \{1, 2, \dots, K-1\} \quad (3.15)$$

Ce modèle est paramétré par une matrice β de taille $(K-1) \times (p+1)$, qui se réduit à un simple vecteur dans le cas d'une classification binaire. Dans sa version multinomiale, le modèle logistique consiste à estimer les probabilités conditionnelles suivantes :

$$P(\mathbf{y} = k | \mathbf{x} = x) = \frac{\exp(\beta_{k,0} + \beta_k^t x)}{1 + \sum_{c=0}^{K-1} \exp(\beta_{c,0} + \beta_c^t x)}, \quad \forall k \in \{1, 2, \dots, K-1\} \quad (3.16)$$

$$P(\mathbf{y} = K | \mathbf{x} = x) = \frac{1}{1 + \sum_{c=0}^{K-1} \exp(\beta_{c,0} + \beta_c^t x)} \quad (3.17)$$

Pour estimer la matrice β , il est courant de chercher à maximiser la log-vraisemblance des probabilités conditionnelles $P(\mathbf{y} = k | \mathbf{x} = x, \beta)$ à l'aide d'algorithmes tels que celui de Newton-Raphson (Hastie et al., 2009).

Faisant partie des approches linéaires et discriminatives, cherchant à modéliser les probabilités *a posteriori* $P(\mathbf{y} = k | \mathbf{x} = x)$, la régression logistique peut être rapprochée d'une autre approche connue de classification : l'analyse discriminante linéaire (LDA) (Hastie et al., 2009). Dans ce cas, les probabilités *a posteriori* sont estimées en supposant que les classes suivent une distribution gaussienne et qu'elles respectent la propriété de variance homoscedastique.

3.1.1.4 Arbre de décision et forêts aléatoires

Les arbres de décision constituent une approche de classification basée sur la définition de séquence de « questions » dont les réponses par « oui ou non » permettent de définir la classe d'appartenance de chaque individu d'un jeu de données. En partant d'un nœud racine, une suite de questions correspondant à des valeurs que peuvent prendre les données est déroulée. Les valeurs des données sont par exemple testées en introduisant des seuils, des intervalles continus ou discrets. La suite des réponses à ces tests définit plusieurs branches qui se terminent par des nœuds feuilles fixant la décision finale quant à la classe d'appartenance d'un individu.

L'avantage d'une telle approche de classification, en plus d'être intuitive et facilement interprétable, est de pouvoir traiter des données qualitatives ou des données pour lesquelles il est difficile de définir une métrique et donc une mesure de proximité entre individus. Elle constitue un moyen efficace de classification lorsque le problème est simple et bénéficie de l'appui d'une connaissance expert. Ce formalisme a par exemple déjà été utilisé pour la classification des tissus urbains à partir des grilles radioconcentriques présentées au chapitre 2, et sera employée dans la suite pour d'autres tâches de

classification basées sur la transposition de connaissances théoriques et qualitatives au chapitre 4.

La construction d'arbres de décision peut être automatisée lorsqu'un jeu de données d'entraînement est disponible pour définir les différentes propriétés discriminant les individus d'un jeu de données dans une tâche de classification. Le cadre de travail CART (*Classification And Regression Tree*) introduit par Breiman et al. (1984) fournit un exemple de procédure générale pour construire des arbres de décision. Cette procédure récursive repose sur des divisions successives du jeu de données en sous-ensembles de plus en plus petits. Chaque nœud de l'arbre représente une décision, c'est-à-dire une division du jeu de données généralement en deux sous-ensembles. On parle alors d'arbre de décision binaire.

Deux critères sont fondamentaux pour construire un arbre de décision (Duda et al., 2000) :

1. *Les propriétés (ou variables) testées à chaque nœud* : que les données soient quantitatives ou qualitatives, une ou plusieurs variables peuvent être testées simultanément pour diviser le jeu de données. L'objectif est de définir à chaque nœud n l'ensemble de p variables qui définira la meilleure coupe selon un critère d'impureté. Ce critère qui peut correspondre à diverses mesures doit permettre d'optimiser la construction des sous-ensembles. En classification, il s'agit d'une mesure d'erreur de classification ou d'entropie qui rend compte du fait que l'arbre est construit de façon à isoler tous les individus d'un même groupe sur une branche. Dans le cas de l'approche CART, l'impureté de Gini² est utilisée (Breiman et al., 1984).
2. *Le critère définissant quand un nœud est considéré comme terminal* : le principe de construction d'un arbre par minimisation d'un critère d'impureté pris sans contrainte entraîne la construction d'un arbre de taille maximale contenant potentiellement autant de nœuds feuilles terminaux qu'il y a d'individus dans le jeu de données. Pour obtenir un arbre moins complexe, il est donc nécessaire de définir un critère d'arrêt à cette méthode de construction récursive. Ce critère peut être imposé en définissant un seuil au-dessous duquel l'impureté d'un groupe est considérée comme suffisamment basse ou stable. Il peut être défini par minimisation d'une erreur de classification par validation croisée ou encore en utilisant un test d'hypothèse.

Plutôt que de définir un critère d'arrêt de la récursion, il est aussi possible de construire l'arbre jusqu'à sa taille maximale et ensuite de l'élaguer. L'idée est de regrouper certaines feuilles qui donnent un même label et qui satisfont un critère de faible augmentation de l'impureté. Cette approche présente l'intérêt d'explorer toutes les divisions possibles et de limiter les « effets horizontaux »

2. L'impureté de Gini $I(t)$ mesure la pureté d'un nœud t via le nombre de points appartenant à des classes différentes : $I(t) = \sum_{i \neq j} p(i|t)p(j|t)$

apparaissant lorsque le critère d'arrêt a été rencontré trop tôt dans la récursion. Si cette approche peut limiter les effets de sur-apprentissage, elle peut vite trouver ses limites en termes de complexité lorsque le jeu d'entraînement est de taille importante.

L'un des principaux inconvénients de la classification par arbre de décision est la sensibilité au jeu de données d'apprentissage : sans élagage, des problèmes de sur-apprentissage peuvent apparaître rapidement ; et le changement d'un simple élément dans les données peut entraîner une erreur de décision qui se répercutera dans tout l'arbre.

Les forêts aléatoires de Breiman (2001) constituent une extension intéressante pour limiter ces effets. Le principe est de construire un modèle prédictif à partir d'un ensemble de T arbres de décision binaire, construits selon l'approche CART enrichie par un double processus de « randomisation ». Chaque arbre $\lambda_t(\mathbf{x}; \mathcal{A})$ est construit à partir d'un échantillon obtenu par *bagging* (ou *bootstrap aggregating*), technique de ré-échantillonnage qui génère à partir d'un échantillon d'entraînement \mathcal{A} de taille m un ensemble $\mathcal{A}^{(t)}$ d'échantillons de taille $l < m$ tirés aléatoirement et avec remise (Breiman, 1996). De plus, un sous-ensemble de variables est tiré aléatoirement pour prendre une décision à chaque nœud de l'arbre, sachant que chaque arbre est de taille maximale (pas d'étape d'élagage). Cette méthode de construction permet de construire un classifieur par agrégation des arbres selon un principe de vote majoritaire. En combinant ces arbres peu corrélés, on obtient un classifieur plus performant qui n'est plus victime des effets du sur-apprentissage.

Les différentes approches et méthodes de classification présentées permettent de traiter des données dans un cadre de travail où celles-ci sont suffisamment bien spécifiées, et suffisamment stables pour assurer la validité du classifieur. Nous allons voir maintenant comment tenir compte de la variabilité ou des changements statistiques qui peuvent faire évoluer un modèle et le rendre invalide pour de nouvelles tâches de classification.

3.1.2 Les approches par transfert d'apprentissage

Les différentes approches de classification présentées jusqu'ici, et plus particulièrement celles se basant sur un cadre d'apprentissage supervisé, supposent que les données d'entraînement et chaque nouvelle donnée à prédire sont tirées de la même distribution. Le jeu de données d'entraînement doit être un sous-ensemble représentatif de l'ensemble du jeu de données à traiter pour obtenir un modèle prédictif de qualité. En effet, si un changement de distribution apparaît entre les données d'entraînement et de test, alors la plupart des modèles prédictifs doivent être reconstruits par rapport au jeu de données test. Or cette hypothèse n'est souvent pas vérifiée en pratique sur des jeux de données réelles, à cause du coût d'acquisition des données labellisées, de la complexité et de la variabilité des données réelles. Bien que cette différence de distribution puisse

entraîner des différences sur les paramètres d'un classifieur, il est souvent trop coûteux, voire même impossible d'acquérir de nouvelles données labellisées pour reconstruire un modèle correct.

La branche du transfert d'apprentissage propose donc une solution pour traiter ce problème en utilisant la connaissance acquise sur une tâche *Source* dans le but de réaliser une tâche *Cible*. Pour introduire les définitions et approches du transfert d'apprentissage nous nous basons sur l'article d'état de l'art de [Pan et Yang \(2010\)](#).

Ce cadre de travail repose sur la définition des notions de *domaine* et de *tâche*. Un domaine, noté \mathcal{D} est constitué de deux composantes : un ensemble de caractéristiques \mathbf{x} (ou covariables pour reprendre le terme précédemment utilisé) et une distribution marginale de probabilité $P(\mathbf{x})$. Étant donné un domaine $\mathcal{D} = \{\mathbf{x}, P(\mathbf{x})\}$ une tâche \mathcal{T} est, elle, constituée de deux composantes : un espace de labels \mathbf{y} et une fonction prédictive $f(\cdot)$ qui sert à prédire le label $y = f(x)$ de tout individu $x \in \mathbf{x}$.

Définition 1 (*Transfert d'apprentissage*)

Étant donnés un domaine *Source* \mathcal{D}_S associé à une tâche *Source* \mathcal{T}_S , et un domaine *Cible* \mathcal{D}_T associé à une tâche *Cible* \mathcal{T}_T , le transfert d'apprentissage a pour but d'améliorer l'apprentissage de la fonction prédictive $f_T(\cdot)$ de \mathcal{D}_T en utilisant la connaissance apprise sur \mathcal{D}_S et \mathcal{T}_S , tout en considérant que $\mathcal{D}_S \neq \mathcal{D}_T$ ou $\mathcal{T}_S \neq \mathcal{T}_T$.

La définition précédente implique que le transfert d'apprentissage regroupe les situations d'apprentissage où l'on cherche à apprendre et/ou améliorer une fonction prédictive f_T en utilisant la connaissance acquise sur un domaine qui est différent soit en termes de covariables, $\mathbf{x}_S \neq \mathbf{x}_T$, soit en termes de distributions marginales, $P(\mathbf{x}_S) \neq P(\mathbf{x}_T)$. Les situations rencontrées peuvent aussi concerner l'utilisation de domaines identiques pour réaliser des tâches d'apprentissages différentes soit par les labels à prédire $\mathbf{y}_S \neq \mathbf{y}_T$, soit par leur distribution $P(\mathbf{y}_S) \neq P(\mathbf{y}_T)$.

Dans la pratique le transfert d'apprentissage peut donc aussi bien concerner des problèmes de classifications supervisées et non supervisées, que des problèmes de régression. Le transfert d'apprentissage trouve son intérêt dans différentes applications, qui concernent par exemple la classification de documents textes ou de pages web ([Do et Ng, 2005](#); [Daumé III et Marcu, 2006](#)), la classification d'images ([Wu et Dietterich, 2004](#)), de données temporelles ([Zheng et al., 2008](#)), ou encore de données biologiques ou socio-économiques ([Bouveyron et Jacques, 2010](#)).

Toutes ces techniques peuvent être réparties en trois cadres de travail, résumés par le tableau 3.1 : le transfert d'apprentissage non supervisé, le transfert d'apprentissage inductif et le transfert d'apprentissage transductif. Le transfert inductif regroupe les situations d'apprentissage où certains labels doivent être disponibles au moins sur \mathcal{D}_T , ce qui correspond aussi à un problème d'apprentissage autodidacte ([Raina et al., 2007](#)) ;

TABLE 3.1 – Différentes approches de transfert d’apprentissage

Approches	Domaines	Tâches	Labels pour \mathcal{T}_S	Labels pour \mathcal{T}_T
<i>Apprentissage classique</i>	$\mathcal{D}_S = \mathcal{D}_T$	$\mathcal{T}_S = \mathcal{T}_T$	Oui/Non	Oui/Non
<i>Transfert inductif</i>	$\mathcal{D}_S = \mathcal{D}_T$	$\mathcal{T}_S \neq \mathcal{T}_T$	Oui/Non	Oui
<i>Transfert non supervisé</i>	$\mathcal{D}_S \neq \mathcal{D}_T$	$\mathcal{T}_S \neq \mathcal{T}_T$	Non	Non
<i>Transfert transductif</i>	$\mathcal{D}_S \neq \mathcal{D}_T$	$\mathcal{T}_S = \mathcal{T}_T$	Oui	Non

et parfois sur \mathcal{D}_S , ce qui définit un problème d’apprentissage multitâche (Caruana, 1998). Le transfert non supervisé, comme le transfert inductif, concerne des problèmes où \mathcal{T}_S et \mathcal{T}_T sont différentes mais reliées. Toutefois, dans le cas non supervisé aucun label n’est disponible sur les deux ensembles de données (Guerra et al., 2010).

Ce dernier cas correspond au problème aussi connu sous le nom de *dataset-shift* (Quionero-Candela et al., 2009), qui regroupe toutes les formes de différences de distribution entre les jeux de données d’apprentissage et de test. Le biais de sélection de l’échantillon est par exemple un problème de *dataset-shift* largement étudié (Vella, 1998). Il intervient lorsque des erreurs lors du processus de sélection et de rejet des individus de l’échantillon entraînent des différences de distribution dépendant de la variable à prédire. Le problème de *covariates shift* (Shimodaira, 2000) correspond au cas plus général où les distributions marginales des covariables diffèrent entre les ensembles d’entraînement et de test.

Remarque 1 (*L’apprentissage transductif*)

Dans le cadre du transfert d’apprentissage transductif, cadre de travail qui nous intéresse ici, les tâches *Source* et *Cible* sont les mêmes. Les covariables peuvent être différentes sur les deux jeux de données, ou bien elles peuvent être identiques mais avec différentes distributions marginales. Ce problème ne doit pas non plus être confondu avec l’apprentissage transductif tel que proposé par Vapnik (2000), qui le définit comme un problème d’apprentissage statistique où les données d’entraînement labellisées et les données de test non labellisées doivent être utilisées simultanément pour prédire les labels des données test. La différence est que même si les covariables sont les mêmes dans le transfert transductif, leurs distributions sont différentes ce qui n’est jamais supposé être le cas en apprentissage transductif.

Dans un problème de transfert transductif appliqué à une tâche de classification \mathcal{T}_T , deux types d’approches sont principalement utilisées pour construire un classifieur adapté à \mathcal{D}_T . La première consiste à chercher un espace de caractéristique commun aux deux domaines, permettant d’étendre le classifieur *Source* $f_S(\cdot)$ au domaine *Cible* (Pan et al., 2008). La seconde consiste à apprendre des liens de distribution entre les domaines *Source* et *Cible* pour pouvoir réutiliser les labels du domaine *Source*. On peut par exemple citer les approches par échantillonnage préférentiel (*importance sampling*)

(Sugiyama et al., 2007), ou les approches par modélisation de liens paramétriques auxquelles nous nous intéresserons plus en détails dans la suite de nos travaux.

On retrouve ces modèles paramétriques dans divers travaux de classification supervisée et non supervisée ou de régression compilés par Beninel et al. (2012), et présentés sur des distributions gaussiennes, de Bernoulli ou logistiques. Cette problématique du transfert d'apprentissage repose sur l'identification d'une application $\phi_k : \mathbb{R}^d \rightarrow \mathbb{R}^d$ liant les distributions des covariables \mathbf{x}_S et \mathbf{x}_T conditionnellement aux labels \mathbf{y} correspondant à K groupes identiques sur les deux tâches :

$$\mathbf{x}_{T|\mathbf{y}_T=k} \sim \phi_k(\mathbf{x}_{S|\mathbf{y}_S=k}) = [\phi_k^1(\mathbf{x}_{S|\mathbf{y}_S=k}), \dots, \phi_k^d(\mathbf{x}_{S|\mathbf{y}_S=k})] \quad (3.18)$$

Deux hypothèses sont classiquement posées sur la forme des fonctions ϕ_k :

1. chaque composante $\phi_k^j(\mathbf{x}_{S|\mathbf{y}_S=k})$ dépend seulement de la j -ème composante de $\mathbf{x}_{T|\mathbf{y}_T=k}$,
2. chaque composante $\phi_k^j(\mathbf{x}_{S|\mathbf{y}_S=k})$ est C^1

Les liens paramétriques considérés sont nécessairement des liens affines ou linéaires de la forme :

$$\mathbf{x}_{T|\mathbf{y}_T=k} \sim D_k \mathbf{x}_{S|\mathbf{y}_S=k} + b_k \quad (3.19)$$

où D_k est une matrice diagonale de $\mathbb{R}^{p \times p}$, et b_k un vecteur de \mathbb{R}^p . Pour réaliser la tâche de classification sur le domaine cible, il est alors nécessaire d'apprendre ces deux paramètres. Différents modèles parcimonieux (Beninel et al., 2012) peuvent être employés et appris grâce à l'algorithme EM. Il s'agira ensuite d'identifier le modèle le plus pertinent pour faire le lien entre les distributions des données.

Jusqu'ici nous nous sommes attachés à présenter des méthodes permettant de caractériser des distributions selon des critères géométriques ou statistiques pour retrouver des groupes de données. Mais nous verrons dans la suite que dans le cas spécifique de la classification de données spatialisées, il est important de prendre en compte d'autres caractéristiques de distribution, celles des structures spatiales.

3.1.3 Les approches spatiales

Qu'elles soient appliquées dans un cadre supervisé ou non supervisé, ou qu'elles cherchent à bénéficier d'un apport de connaissance par transfert d'apprentissage, les approches de classification précédentes peuvent être vues comme des approches généralistes utilisables dans différents domaines d'application. Dans le cas probabiliste, ces approches reposent sur une hypothèse générale qui veut que les données soient représentées par des variables i.i.d.. Lorsque que la tâche de classification concerne des données spatialisées cette hypothèse devient problématique puisqu'une des hypothèses fondamentales de l'analyse spatiale est au contraire de supposer que des entités proches

s'influencent mutuellement : les données ne peuvent donc pas être considérées comme indépendantes.

Nous nous intéressons donc ici aux méthodes de classification permettant de tenir compte de ces influences locales induisant de l'auto-corrélation spatiale, et en particulier aux méthodes issues du traitement d'image. Nous distinguerons les approches de segmentation des approches de classification par pixel. Une image segmentée est une image où chaque pixel est affecté à un groupe homogène et connexe. La segmentation produit des *régions d'image* représentant toutes des groupes distincts. La classification par pixel produit, elle, des images classées où chaque pixel est affecté à un groupe homogène qui n'est pas forcément connexe. Les groupes définissent dans ce cas des classes qui sont homogènes statistiquement, mais qui ne le sont pas nécessairement spatialement.

Nous introduirons brièvement les approches de la segmentation d'images. Nous nous intéresserons surtout aux méthodes de classification par-pixel, qui seront utilisées pour produire des classifications des tissus urbains. Nous nous focaliserons sur le formalisme des champs de Markov, et nous introduirons différents algorithmes de classification markovienne.

3.1.3.1 Les approches de segmentation d'image

Les approches de segmentation d'image sont nombreuses et se retrouvent dans des domaines d'application variés. Ici nous nous intéresserons aux approches fournissant une partition complète de l'image, où chaque pixel considéré en entrée est affecté à un groupe en sortie.

Les approches régions Les approches région regroupent les méthodes de segmentation qui cherchent directement à construire des régions d'images. Deux formalismes sont envisageables : les algorithmes par croissance de régions basés sur le regroupement de pixels autour de certains pixels *seeds*, et les algorithmes par partage de régions basés sur la décomposition séquentielle de l'image en plusieurs régions.

Parmi les algorithmes de croissance de régions, on retrouve l'algorithme de la ligne de partage des eaux (LPE) développé dans le cadre de la morphologie mathématique (Meyer et Beucher, 1990) ou les algorithmes de clustering par estimation de densité par noyau tels que l'algorithme *Mean-Shift* (Cheng, 1995). On retrouve aussi dans ce cadre les algorithmes de segmentation par partitionnement hiérarchique du graphe associé à l'image utilisant des critères de coupe tels que les *normalized cuts* de Shi et Malik (2000).

Dans la catégorie du partage de régions l'approche la plus simple est la segmentation par seuillage. Les méthodes pour identifier ces seuils reposent sur l'étude des

caractéristiques globales de l'image, notamment par la détection de modes sur l'histogramme d'une image (Sonka et al., 2008). On retrouve aussi dans cette catégorie des algorithmes de type *Split and Merge* combinant les deux formalismes du partage et de la croissance de régions pour construire une représentation pyramidale de l'image.

Approche frontière Les méthodes de segmentation selon l'approche frontière reposent sur l'identification de régions non pas par les pixels qui les composent, mais par la recherche des frontières entre régions. Ces algorithmes reposent sur deux phases de traitement : la détection de contours saillant dans l'image ; la segmentation de l'image selon les contours détectés. La phase la plus importante est donc la phase de détection de ces contours, sachant qu'il existe une variété de filtres pour faire de la détection de contours. Les plus simples reposent sur différents calculs de gradients comme pour les filtres de Canny, de Sobel ou de Prewitt (Sonka et al., 2008).

Pour gérer le bruit, il est souvent nécessaire de reconnecter les contours significatifs fragmentés par des méthodes locales ou globales. Certains filtres plus complexes reposent sur la modélisation de processus de diffusion anisotropique (Perona et Malik, 1990), sur la transformée de Hough (Ballard, 1981) ou encore sur l'utilisation d'EDP et sur l'équation de la chaleur (Catté et al., 1992). Diverses approches peuvent ensuite être employées pour produire la segmentation finale, sachant que le problème le plus courant est de distinguer les contours les plus pertinents parmi un ensemble de contours bruités.

3.1.3.2 Modélisation par champs de Markov

Le formalisme markovien, développé initialement dans le domaine de la physique statistique (Chandler, 1987), fournit un cadre de travail adapté à la prise en compte de toute sorte de structure de dépendances entre les individus d'un jeu de données. Largement employé en traitement d'image et de données spatialisées, il permet de prendre en compte l'autocorrélation spatiale introduite par les relations de dépendances statistiques entre pixels voisins. Si les modèles markoviens sont employés dans diverses tâches de traitement d'images relatives à la restauration, la reconnaissance de formes ou encore l'analyse de textures (Pieczynski, 2003; Li, 2009), nous nous intéresserons ici aux tâches de classification par pixel.

Soit I une image composée d'un ensemble de n pixels, aussi appelés *sites*. Pour modéliser la structure de dépendances spatiales de l'image, il convient de définir pour chaque pixel $i \in I$ un voisinage noté $N(i)$ composé par exemple des 4 ou 8 plus proches voisins de i . On note $\mathcal{N}(I)$ le système de voisinage regroupant tous les voisinages $N(i)$, et on abrège une relation de voisinage $j \in N(i)$ à la notation $i \sim j$ puisque tous les voisinages seront considérés comme étant symétriques : $j \in N(i) \iff i \in N(j)$.

Définition 2 (*Clique*)

Un sous-ensemble c de I est appelé clique relative au système de voisinage $\mathcal{N}(I)$, si c est un singleton ou un ensemble de pixels mutuellement voisins de $N(i)$. Une clique est dite d'ordre r si elle contient r pixels.

Pour des tâches de classification par pixel, on considère souvent les cliques d'ordre un, c'est-à-dire les singletons, et les cliques d'ordre deux composées de paires de pixels voisins (figure 3.1).

Définition 3 (*Champ de Markov*)

Une collection de variables aléatoires discrètes $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ définies sur I est un champ de Markov si la distribution de probabilité $P(\mathbf{y})$ est définie selon les propriétés suivantes :

$$\forall i, \quad P(y_i | \mathbf{y}_{I \setminus \{i\}}) = P(y_i | y_j, j \in N(i)), \quad (3.20)$$

$$P(\mathbf{y}) > 0. \quad (3.21)$$

La propriété (3.20) indique que seuls les voisins de i influencent la valeur du champ au point i , tandis que la propriété (3.21) assure que $P(\mathbf{y})$ est définie de manière unique. Pour avoir une expression plus exploitable d'un champ de Markov, il est courant d'utiliser le théorème Hammersley-Clifford qui permet de définir un champ de Markov selon une distribution jointe (Besag, 1974).

Définition 4 (*Distribution de Gibbs*)

Un champ aléatoire \mathbf{y} est un champ de Markov si \mathbf{y} a une distribution de Gibbs de la forme :

$$P(\mathbf{y}) = W^{-1} \exp(-H(\mathbf{y})) \quad (3.22)$$

où H est la fonction énergie, définie comme une somme de fonctions potentielles V_c associées à des cliques c :

$$H(\mathbf{y}) = \sum_{c \in C} V_c(\mathbf{y}_c) \quad (3.23)$$

et W est la constante de normalisation aussi appelée fonction de partition telle que :

$$W = \sum_{\mathbf{y}} \exp(-H(\mathbf{y})) \quad (3.24)$$

Puisque le calcul de la constante de normalisation W implique toutes les réalisations de \mathbf{y} , il n'est pas possible de calculer en pratique la distribution jointe $P(\mathbf{y})$ sous une forme exacte. Il est cependant possible de calculer les probabilités conditionnelles :

$$P(y_i | \mathbf{y}_{N(i)}) = \frac{\exp(-\sum_{c \ni i} V_c(\mathbf{y}_c))}{\sum_{y_i} \exp(-\sum_{c \ni i} V_c(\mathbf{y}_c))}, \quad (3.25)$$

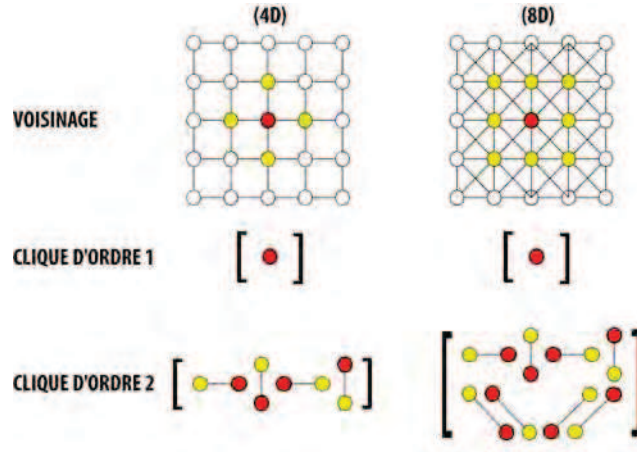


FIGURE 3.1 – Systèmes de voisinages 4-D et 8-D et cliques associées

En se basant sur cette expression, [Besag \(1975\)](#) a proposé une approximation de la vraisemblance d'un champ de Markov, connue sous le nom de pseudo-vraisemblance :

$$\mathcal{PL}(\mathbf{y}) = \prod_{i \in I} P(y_i | \mathbf{y}_{N(i)}) \quad (3.26)$$

Modèles de Markov discrets Pour modéliser un champ de Markov associé à une image, il est nécessaire de définir la structure spatiale associée à ce champ. Nous donnons ici des exemples de modèles communément employés pour représenter une variable discrète \mathbf{y} , fournissant généralement le label associé à chaque pixel $i \in I$ dans une tâche de classification. Ces modèles sont généralement définis à partir de leur fonction énergie H , donnée en fonction des fonctions potentielles différenciées selon l'ordre des cliques.

Le modèle markovien le plus ancien est le modèle d'[Ising \(1925\)](#) développé en physique statistique et correspondant au cas où la variable discrète \mathbf{y} ne peut prendre que les deux valeurs -1 et 1 :

$$H(\mathbf{y}) = \alpha \sum_i y_i - \beta \sum_{i \sim j} y_i^t y_j \quad (3.27)$$

Dans ce modèle, les paramètres α et β sont des constantes à valeurs réelles représentant un champ magnétique externe et une constante d'interactions entre des sites voisins.

Le modèle de Potts ([Sigelle et Ronfard, 1992](#)) constitue une extension du modèle d'Ising au cas où la variable \mathbf{y} prend ses valeurs parmi $[0,1]^K$. Il est défini à partir des cliques d'ordre 2 et peut s'écrire :

$$H(\mathbf{y}) = - \sum_{i \sim j} y_i^t \beta y_j \quad (3.28)$$

Dans la version la plus simple du modèle, $\beta = b\mathbb{I}$ où b est une constante modélisant les interactions entre classes de façon homogène. Il est cependant possible de définir β comme une matrice symétrique de taille $K \times K$ modélisant des interactions de forces

différentes selon les classes. Employer cette matrice plutôt qu'une simple constante d'interaction peut s'avérer utile pour la classification lorsque les classes ont des structures spatiales différentes, bien que cela complique l'estimation du modèle.

Ce modèle de Potts peut aussi être étendu à une version avec un champ externe $\alpha \in \mathbb{R}^K$:

$$H(\mathbf{y}) = - \sum_i \alpha y_i - \sum_{i \sim j} y_i^t \beta y_j \quad (3.29)$$

L'intégration de ce paramètre peut s'avérer intéressante pour des cas de classifications où les classes ont des fréquences d'apparitions très différentes (Celeux et al., 2004).

Les champs de Markov cachés peuvent être utilisés pour réaliser des tâches de classification par pixel en reprenant le formalisme des variables cachées. La variable de classe \mathbf{y} est modélisée par un champ de Markov discret caché, tandis que les covariables \mathbf{x} sont modélisées par une densité de probabilité f paramétrée par $\theta = (\theta_1, \dots, \theta_K)$ puisque les variables sont supposées conditionnellement indépendantes à \mathbf{y} . La distribution jointe du couple (\mathbf{x}, \mathbf{y}) peut alors s'écrire :

$$P(\mathbf{x}, \mathbf{y} | \psi) = f(\mathbf{x} | \mathbf{y}, \theta) P(\mathbf{y} | \beta) \quad (3.30)$$

$$= \prod_{i \in I} f(x_i | y_i, \theta) W(\beta)^{-1} \exp(-H(\mathbf{y} | \beta)) \quad (3.31)$$

$$= W(\beta)^{-1} \exp \left(-H(\mathbf{y} | \beta) + \sum_{i \in I} \log f(x_i | y_i, \theta) \right) \quad (3.32)$$

où $\psi = (\beta, \theta_1, \theta_2, \dots, \theta_K)$, et β est le paramètre de la fonction énergie H du champ de Markov \mathbf{y} . De l'équation (3.32) il découle que le champ conditionnel $\mathbf{y} | \mathbf{x}$ est aussi un champ de Markov avec une fonction d'énergie de la forme :

$$H(\mathbf{y} | \mathbf{x}, \psi) = H(\mathbf{y} | \beta) - \sum_{i \in I} \log f(x_i | y_i, \theta) \quad (3.33)$$

Pour une tâche de classification, on cherche alors à estimer la distribution $P(\mathbf{y} | \mathbf{x})$ pour déterminer les labels de classes avec par exemple la règle du MAP. Cette approche de classification est semblable à celle présentée pour les modèles de mélanges dans la section 3.1.1.1, à la différence que l'estimation des paramètres sera complexifiée par le fait que le mélange n'est pas indépendant. Dans la pratique deux approches de classification sont alors utilisées : celles qui supposent les paramètres $\psi = (\beta, \theta)$ connus et celles qui cherchent à les estimer.

3.1.3.3 Algorithmes de classification à paramètres connus

Algorithme ICM Lorsque les paramètres sont supposés connus, l'estimation du MAP peut se faire à l'aide d'algorithmes tels que l'algorithme ICM (*Iterative conditional Mode*) introduit par Besag (1986). L'algorithme consiste à estimer le MAP selon

une procédure itérative reposant sur la mise à jour séquentielle des y_i à partir des valeurs des covariables x_i et de la configuration courante de \mathbf{y} . A l'itération q , le label $y_i^{(q)}$ d'un seul pixel i est mis à jour par :

$$y_i^{(q)} = \arg \max_{y_i} P(y_i | \mathbf{y}_{N(i)} = \mathbf{y}_{N(i)}^{(q-1)}, \mathbf{x}), \quad (3.34)$$

sachant que $\mathbf{y}^{(q-1)}$ représente la configuration courante fixée pour tout pixel $j \neq i$. Ce calcul itératif de l'estimateur est facilité par le fait que seules des probabilités locales, qui ne font intervenir que les K valeurs possibles de y_i sont calculées. Un cycle de l'algorithme est effectué lorsque tous les pixels de l'image ont été mis à jour. En pratique, on observe qu'après quelques cycles cet algorithme glouton converge vers un maximum local de la distribution de Gibbs du champ conditionnel $\mathbf{y}|\mathbf{x}$. Le résultat de la procédure est toutefois très dépendant de l'initialisation $\mathbf{y}^{(0)}$.

Algorithme du recuit simulé Le MAP peut aussi être calculé à l'aide d'un algorithme tel que le recuit simulé de [Geman et Geman \(1984\)](#). Cet algorithme repose sur une méthode de relaxation stochastique qui correspond à la recherche d'une configuration d'énergie minimale d'une distribution de Gibbs. Le principe est d'introduire une distribution de probabilité conditionnelle dépendant d'un paramètre de température $\mathcal{T} > 0$:

$$P_{\mathcal{T}}(\mathbf{y}|\mathbf{x}) = P(\mathbf{y}|\mathbf{x})^{1/\mathcal{T}} \propto \exp\left(-\frac{H(\mathbf{y}|\mathbf{x})}{\mathcal{T}}\right) \quad (3.35)$$

Lorsque $\mathcal{T} \rightarrow \infty$, $P_{\mathcal{T}}(\mathbf{y}|\mathbf{x})$ converge vers une distribution uniforme. Mais lorsque $\mathcal{T} \rightarrow 0$, $P_{\mathcal{T}}(\mathbf{y}|\mathbf{x})$ converge vers $P(\mathbf{y}|\mathbf{x})$. Le principe de l'algorithme est donc de simuler une suite de configuration de la distribution $P_{\mathcal{T}}$ en faisant décroître progressivement la température \mathcal{T} afin de converger vers le maximum global. L'algorithme peut se résumer de la façon suivante :

- Initialisation d'une configuration $\mathbf{y}^{(0)}$ et d'une température $\mathcal{T}^{(0)}$
- A l'itération q : simulation de la configuration $\mathbf{y}^{(q)}$ selon la loi $P_{\mathcal{T}^{(q)}}(\mathbf{y}^{(q)}|\mathbf{x})$ à l'aide de l'échantillonneur de Gibbs ou de l'algorithme de Metropolis-Hastings ([Hastings, 1970](#))
- Diminution lente de la température : $\mathcal{T}^{(q)} > \frac{c}{\log(1+q)}$, où c est une constante ([Geman et Geman, 1984](#)), jusqu'à ce qu'il y ait convergence ou qu'un critère d'arrêt soit rencontré.

La convergence de l'algorithme dépend de la vitesse de décroissance de la température \mathcal{T} , sachant qu'une décroissance trop rapide peut entraîner la procédure dans un maximum local mais qu'une décroissance trop lente implique une convergence très lente.

Échantillonneur de Gibbs Pour mettre en œuvre l'algorithme du recuit simulé et pour réaliser d'autres tâches impliquant notamment l'estimation de paramètres, il est

courant d'avoir recours à l'échantillonneur de Gibbs proposé par **Geman et Geman (1984)** pour simuler des réalisations d'un champ de Markov \mathbf{y} . Cet algorithme qui est un cas particulier de l'algorithme de Metropolis-Hastings, est une méthode de type Chaîne de Markov de Monte-Carlo (MCMC). Une procédure itérative est utilisée pour simuler une réalisation de la distribution de \mathbf{y} , en se basant ici aussi uniquement sur les caractéristiques locales des y_i .

En partant d'une configuration initiale $y_1^{(0)}, \dots, y_n^{(0)}$, l'algorithme consiste à générer à chaque itération q , une réalisation de $\mathbf{y}^{(q)}$, où un seul pixel i est visité. Deux réalisations $\mathbf{y}^{(q-1)}$ et $\mathbf{y}^{(q)}$ ne peuvent différer que par la valeur y_i du pixel i visité à l'itération q . A l'itération q la simulation de \mathbf{y} se fait selon la procédure suivante :

- Tirer uniformément un pixel i parmi l'ensemble des pixels de l'image.
- Tirer y_i selon la loi conditionnelle locale : $P(y_i | \mathbf{y}_{N(i)}^{(q-1)}, \mathbf{x})$

En pratique cet algorithme génère une suite de q images. On considère que l'algorithme a convergé après un grand nombre d'itération ou lorsqu'il est stable entre deux itérations. Si l'algorithme peut être un peu long à converger puisque les pixels sont mis à jour un à un, il présente toutefois l'avantage d'être facile à mettre en œuvre puisqu'il ne nécessite pas le calcul de la fonction de partition.

Approximation en champ moyen L'approximation en champ moyen qui vient elle aussi de la physique statistique (**Chandler, 1987**) est utilisée dans diverses tâches d'apprentissage statistique puisqu'elle permet d'approximer la distribution d'un champ de Markov \mathbf{y} . Le principe de la théorie du champ moyen est de fixer les interactions entre un pixel $i \in I$ et ses voisins $j \in N(i)$, de sorte à ce que l'on puisse considérer que chaque pixel interagit seulement avec son voisinage. Cela revient à négliger les interactions entre un pixel i et les pixels hors de son voisinage $N(i)$ et à résumer les interactions avec les pixels au sein de son voisinage à une valeur moyenne. Cette procédure a pour objectif de transformer l'ensemble des variables aléatoires $\mathbf{y} = (y_1, \dots, y_n)$ en un système composé de variables indépendantes. La distribution marginale d'un champ de Markov \mathbf{y} peut alors être approximée de la façon suivante :

$$P(\mathbf{y}) \simeq P_{\mathbf{y}^{mf}}(\mathbf{y}) = \prod_{i \in I} P(y_i | \mathbf{y}_{N(i)}^{mf}) \quad (3.36)$$

où \mathbf{y}^{mf} est le champ voisin qui est une collection de valeurs connues $\{y_i^{mf}, \forall i \in I\}$, correspondant aux valeurs de l'espérance par rapport à la loi de $P_{\mathbf{y}^{mf}}$.

Lorsque le champ de Markov \mathbf{y} est défini selon un modèle de Potts à K classes (équation (3.28)), la distribution approximée $P_{\mathbf{y}^{mf}}(y_i | \beta)$ peut être écrite pour tout $i \in I$:

$$P_{\mathbf{y}^{mf}}(y_i | \beta) = W^{mf-1} \exp[-H^{mf}(y_i | \beta)] \quad (3.37)$$

$$\text{et } H^{mf}(y_i | \beta) = - \sum_{j \sim i} y_i^t \beta y_j^{mf} \quad (3.38)$$

où $H^{mf}(y_i|\beta)$ est l'énergie en champ moyen locale, et W^{mf} est la fonction de partition locale associée. Pour calculer cette distribution pour tout $i \in I$ il est nécessaire de connaître les valeurs du champ voisin pour chaque $j \in N(i)$, ce qui revient à calculer les $\{y_i^{mf}\}$ pour toute l'image. Pour un modèle de Potts on doit alors calculer pour tout $i \in I$:

$$y_i^{mf} = \frac{\sum_{y_i} y_i \exp(y_i^t \beta \sum_{j \sim i} y_j^{mf})}{\sum_{y_i} \exp(y_i^t \beta \sum_{j \sim i} y_j^{mf})} \quad (3.39)$$

Le calcul de ce champ voisin revient donc à résoudre une série de n équations. L'estimation des y_i^{mf} peut se faire de façon séquentielle ou parallèle puisque le problème repose sur des calculs locaux (Zhang, 1992).

En supposant les paramètres du champ de Markov conditionnel $\mathbf{y}|\mathbf{x}$ connus, la procédure de classification par la règle du MAP revient alors à calculer l'estimateur sur la distribution approximée :

$$\forall i \in I, \quad y_i^{MAP} = \arg \max_{y_i} P(y_i | \mathbf{y}_{N(i)}) = \mathbf{y}_{N(i)}^{mf}, x_i \quad (3.40)$$

3.1.3.4 Approche mélange pour les champs de Markov

Les algorithmes présentés précédemment demandent de connaître les paramètres du modèle markovien, ce qui n'est généralement pas le cas en pratique. Un grand nombre de méthodes ont donc été proposées pour résoudre le problème d'estimation des paramètres d'un champ de Markov, problème complexifié par la présence des dépendances spatiales et le calcul de la fonction de partition. Plusieurs solutions reposent sur des simulations de Monte-Carlo comme l'algorithme du gradient stochastique de Younes (1988), l'EM Gibbsien de Chalmond (1989), ou encore l'algorithme MCEM de Wei et Tanner (1990). D'autres reposent sur l'utilisation d'approches variationnelles et notamment sur l'approximation en champ moyen comme l'algorithme de Zhang (1992) ou sur l'algorithme NREM de Celeux et al. (2003). L'utilisation de la pseudo-vraisemblance de Besag (1974) (équation (3.26)) constitue une autre approche pour contourner la complexité du problème et intégrer une étape d'estimation dans l'algorithme ICM par exemple (Besag, 1986). On peut aussi citer la procédure d'estimation de paramètres plus générale de l'algorithme ICE de Pieczynski (1994).

L'algorithme NREM Nous nous intéresserons plus en détail à l'algorithme NREM introduit par Celeux et al. (2003). Cette version de l'algorithme EM basée sur une approximation de type champ moyen permet de faire de l'estimation des paramètres d'un champ de Markov pour des tâches de classification par pixel non supervisée dans le cadres de modèles de mélanges non indépendants. Pour pouvoir se ramener à un modèle de mélange indépendant sur lequel on puisse appliquer l'algorithme EM, il est proposé d'utiliser une approximation en champ moyen de la distribution $P(\mathbf{x}, \mathbf{y}|\beta)$:

$$P_{\mathbf{y}^{mf}}(\mathbf{x}, \mathbf{y}|\beta) = \prod_i P_{\mathbf{y}^{mf}}(x_i, y_i|\beta) = \prod_i P(y_i | \mathbf{y}_{N(i)}^{mf}) f(x_i; \theta_{y_i}) \quad (3.41)$$

L'algorithme NREM se présente ainsi comme une procédure itérative alternant une phase d'estimation du champ voisin \mathbf{y}^{mf} , notée NR pour *Neighborhood Restoration*, et une phase d'estimation des paramètres du champ de Markov par l'algorithme EM applicable sur le mélange indépendant défini sous l'approximation en champ moyen. En partant d'une initialisation des paramètres du modèle $\psi^{(0)} = (\theta^{(0)}, \beta^{(0)})$, l'algorithme est donné à l'itération q par :

(NR) A partir des paramètres courants $\psi^{(q)}$, calculer le champ voisin $\mathbf{y}^{mf(q)}$ à partir de la distribution conditionnelle approximée :

$$P_{\mathbf{y}^{mf}}(\mathbf{y}|\mathbf{x}, \psi^{(q)}) = \prod_{i \in I} P(y_i | x_i, \mathbf{y}_{N(i)}^{mf}, \psi^{(q)}) \quad (3.42)$$

dont la fonction énergie est donnée par :

$$H^{mf}(\mathbf{y}|\mathbf{x}, \psi) = H^{mf}(\mathbf{y}|\beta) - \sum_{i \in I} \log f(x_i | y_i; \theta) \quad (3.43)$$

(EM) Estimer les paramètres par une itération de l'algorithme EM, qui revient ici à maximiser l'espérance de la log-vraisemblance approximée par :

$$\mathcal{Q}(\psi|\psi^{(q)}) = \sum_{i \in I} \sum_{k \in \mathcal{K}} t_{ik}^{(q)} \log[\pi_{ik}^{(q)} f(x_i; \theta_k^{(q)})] \quad (3.44)$$

où $t_{ik}^{(q)}$ désigne les probabilités *a posteriori*. La procédure se décompose alors en :

(E) Estimer les $t_{ik}^{(q)}$ données par :

$$t_{ik}^{(q)} = \frac{\pi_{ik}^{(q)} f(x_i; \theta_k^{(q)})}{\sum_l \pi_{il}^{(q)} f(x_i; \theta_l^{(q)})}, \quad (3.45)$$

avec $\pi_{ik}^{(q)} = P(y_i | x_i, \mathbf{y}_{N(i)}^{mf}, \psi^{(q)})$.

(M) Mettre à jour les paramètres $\psi^{(q+1)}$ selon les formules :

$$\theta^{(q+1)} = \arg \max_{\theta} \sum_i \sum_k t_{ik}^{(q)} \log f(x_i; \theta_k) \quad (3.46)$$

$$\beta^{(q+1)} = \arg \max_{\beta} \sum_i \sum_k t_{ik}^{(q)} \log \pi_{ik} \quad (3.47)$$

Pour la réalisation de cette procédure [Celeux et al. \(2003\)](#) font remarquer deux points concernant l'étape NR. D'une part il est possible et préférable d'utiliser la distribution conditionnelle $P_{\mathbf{y}^{mf}}(\mathbf{y}|\mathbf{x})$ plutôt que la distribution $P_{\mathbf{y}^{mf}}(\mathbf{y})$. Cela permet d'obtenir une restauration plus précise du champ voisin \mathbf{y}^{mf} basée sur la prise en compte des covariables \mathbf{x} . D'autre part il est possible de fixer les valeurs du champ voisin à des constantes autres que l'espérance de la loi de $P_{\mathbf{y}^{mf}}$, tant que l'équation (3.36) est vérifiée. Ils proposent alors les trois procédures suivantes pour réaliser l'étape NR :

1. **Algorithme en champ moyen** : Les valeurs du champ voisin y_i^{mf} sont fixées aux valeurs moyennes en utilisant l'algorithme de l'approximation en champ moyen.
2. **Algorithme en champ modal** : Les valeurs du champ voisin y_i^{mf} sont fixées aux valeurs modales en utilisant l'algorithme ICM.
3. **Algorithme en champ simulé** : Les valeurs du champ voisin y_i^{mf} sont fixées à une réalisation de la distribution de Gibbs de $P(\mathbf{y}|\mathbf{x})$ obtenue par l'échantillonneur de Gibbs.

Comme pour l'algorithme EM, cet algorithme est influencé par l'initialisation choisie. Les paramètres d'initialisations peuvent par exemple être choisis en effectuant r lancés de l'algorithme NREM à une itération et en choisissant les paramètres du lancé qui maximise la vraisemblance. De plus, les trois procédures employées pour calculer le champ voisin aboutiront à trois solutions localement optimales différentes qui peuvent influencer les sorties de l'algorithme. L'aspect variationnel de l'algorithme en champ moyen garantit que le champ converge vers un optimum local (Wainwright et Jordan, 2008). Le champ moyen fournit alors une solution qui assure que la distribution $P_{\mathbf{y}^{mf}}(\mathbf{y}|\mathbf{x},\psi)$ est la meilleure approximation de $P(\mathbf{y}|\mathbf{x},\psi)$ au sens de la divergence de Kullback-Leibler³. Avec la procédure en champ modal, Celeux et al. (2003) indique que le champ \mathbf{y}^{mf} converge vers un maximum local de $P(\mathbf{y}|\mathbf{x},\psi)$, mais la bonne approximation de $P_{\mathbf{y}^{mf}}(\mathbf{y}|\mathbf{x},\psi)$ est plus incertaine. Pour la procédure en champ simulé, Peyrard (2001) montre des preuves de la convergence théorique de l'algorithme. Cette version de l'algorithme proche d'un algorithme EM stochastique présente l'avantage de fournir une réalisation de la distribution $P(\mathbf{y}|\mathbf{x},\beta)$ qui devient de plus en plus précise lorsque les itérations augmentent et convergent vers les vrais paramètres. Dans la pratique l'algorithme est lancé sur un grand nombre d'itérations et arrêté lorsque l'approximation du champ voisin ne bouge plus entre deux itérations. La classification est ensuite obtenue en appliquant la règle du MAP.

L'algorithme NREM permet donc d'estimer tous les paramètres d'un champ de Markov conditionnel. Il permet d'obtenir une image classée dont les classes peuvent être spécifiées précisément d'un point de vue statistique et spatial. Il offre une base de travail que nous pourrions adapter pour répondre à nos différents objectifs. La mise au point d'une méthode de classification prenant en compte des dépendances spatiales, applicables dans un cadre supervisé mais avec une connaissance minimale des images à classer, pourra donc s'appuyer sur cette méthode d'estimation de paramètres qui doit être revue dans un cadre d'apprentissage par transfert transductif.

3. La divergence de Kullback-Leiber est une mesure de dissimilarité entre deux distributions P et Q : $KL(Q||P) = \sum_{\mathbf{y}} Q(\mathbf{y}) \log \frac{Q(\mathbf{y})}{P(\mathbf{y})}$

3.2 Transfert d'apprentissage et champ de Markov

Dans des domaines liés aux données spatialisées, les méthodes de transfert d'apprentissage trouvent un intérêt particulier puisqu'un simple changement de localisation peut entraîner un changement dans la structure spatiale ou dans la distribution statistique du phénomène observé. L'application d'algorithmes de classification supervisée se trouve donc entravée par ces phénomènes d'hétérogénéité spatiale. Les méthodes de transfert transductif peuvent être mobilisées dans ce contexte pour résoudre ce problème. Dans ce cadre, la connaissance des tissus urbains d'une aire urbaine peut être réutilisée pour classer les tissus urbains d'une autre aire urbaine, sans avoir recours à la construction d'une nouvelle base d'apprentissage qui doit être statistiquement et spatialement représentative de cette aire urbaine.

Dans cette section, nous proposons un algorithme de classification par transfert d'apprentissage transductif adapté au traitement de données images ou de tout type de données à dépendances spatiales. Le problème, qui peut être généralisé, fait écho à nos objectifs de construction d'une base de données des tissus urbains sur l'ensemble des aires urbaines françaises. Il correspond à un problème de classification supervisée d'une collection d'images multivariées, sachant qu'une seule image dont les pixels sont labellisés est disponible, et que les images *Cibles* de la collection peuvent avoir une distribution conditionnelle et/ou une structure spatiale par classe différente de l'image *Source*. Nous traiterons donc le cas de transfert transductif d'une image *Source* aux caractéristiques connues à une image *Cible* que nous cherchons à classer.

3.2.1 Modèles de transfert dans le cadre Markovien

En reprenant le cadre du transfert d'apprentissage (cf. section 3.1.2), on suppose que les données consistent en deux images multivariées I_S et I_T . L'image I_S est associée au domaine *Source* sur lequel il est possible de réaliser une tâche de classification supervisée à partir des labels disponibles. L'image I_T est, elle, associée au domaine *Cible* ne possédant pas de labels. Le but est d'utiliser le modèle markovien facilement appris sur I_S pour en déduire un modèle markovien adapté à I_T .

3.2.1.1 Contraintes sur les covariables

Rappelons que pour réaliser une tâche de classification dans le cadre markovien, l'objectif est d'estimer la distribution $P(\mathbf{y}|\mathbf{x})$ du champ conditionnel $\mathbf{y}|\mathbf{x}$ dont la fonction énergie est donnée par l'équation (3.33). On suppose ici que les covariables \mathbf{x}_S associées à l'image I_S suivent une distribution gaussienne de la forme $\mathbf{x}_{S|\mathbf{y}_S=k} \sim \mathcal{N}(\mu_k^S, \Sigma_k^S)$ aux paramètres connus $\theta^S = (\mu_1^S, \Sigma_1^S, \dots, \mu_K^S, \Sigma_K^S)$. Les covariables \mathbf{x}_T associées à l'image à classer I_T sont, elles, supposées suivre une autre distribution

gaussienne $\mathbf{x}_{T|\mathbf{y}_T=k} \sim \mathcal{N}(\mu_k^T, \Sigma_k^T)$ dont les paramètres θ^T ne sont pas connus.

Pour réaliser le transfert d'apprentissage, on suppose qu'il est possible d'apprendre un lien entre les distributions conditionnelles $\mathbf{x}_S|\mathbf{y}_S$ et $\mathbf{x}_T|\mathbf{y}_T$ qui correspond à un lien paramétrique linéaire de la forme :

$$\mathbf{x}_{T|\mathbf{y}_T=k} \sim D_k \mathbf{x}_{S|\mathbf{y}_S=k} \quad (3.48)$$

En exprimant ces liens linéaires par rapport aux paramètres de distributions gaussiennes, les liens entre distributions deviennent :

$$\mu_k^T = D_k \mu_k^S \quad \text{et} \quad \Sigma_k^T = D_k \Sigma_k^S D_k \quad (3.49)$$

L'équation (3.49) implique que l'estimation des paramètres de la distribution de $\mathbf{x}_{T|\mathbf{y}_T=k}$ se réduit à l'estimation des matrices $D_k, \forall k \in \{1, \dots, K\}$. Sur ces matrices qui expriment la distorsion entre les distributions, il semble alors naturel d'imposer différents types de contraintes reflétant différents types de liens statistiques. [Biernacki et al. \(2002\)](#) ont proposé cinq types de contraintes sur les matrices D_k que nous reprenons ici :

- (M_1) : $D_k = \mathbb{I}_d$, les distributions des covariables sont identiques pour chaque classe k .
- (M_2) : $D_k = \alpha \mathbb{I}_d$, la distorsion est indépendante des covariables et des classes.
- (M_3) : $D_k = D$, la distorsion dépend seulement des covariables.
- (M_4) : $D_k = \alpha_k \mathbb{I}_d$, la distorsion dépend seulement des classes.
- (M_5) : $D_k = D_k$, la distorsion dépend des covariables et des classes (cas général).

Ces modèles correspondent à différentes situations de mise à l'échelle des paramètres θ_T qui ajusteront les centres de classes et leurs variances. Le modèle $M1$ correspond à la situation la plus simple, où l'apprentissage supervisé du modèle *Source* est appliqué sans adaptation au domaine *Cible*. Le modèle $M2$ correspond à la situation où toutes les classes et variables sont ajustées de la même façon, alors que le modèle $M3$ correspond à la situation où toutes les classes sont ajustées de la même façon mais en tenant compte de spécificités de distribution des variables. Ces deux modèles préservent la propriété d'homoscédasticité lorsqu'elle est présente dans le mélange *Source*. Le modèle $M4$ correspond à la situation où chaque classe est ajustée indépendamment mais de manière identique pour chaque variable. Et enfin le modèle $M5$ correspond à la situation générale et plus complexe où la distribution du modèle *Cible* doit être ajustée en fonction de chaque classe et de chaque variable.

3.2.1.2 Contraintes sur le paramètre spatial

Afin de modéliser la distribution spatiale des labels, le champ de Markov caché \mathbf{y} est défini avec un modèle de Potts simple (équation (3.28)) ou avec un champ externe (équation (3.29)) où la matrice d'interactions β de taille $K \times K$ est supposée symétrique.

Même si l'estimation d'une telle matrice est plus compliquée que la simple forme $\beta = b\mathbb{I}$, cette version donnera plus de flexibilité aux structures d'interactions entre classes qui peuvent être très différentes d'une classe à l'autre, et d'une image à l'autre.

Dans les travaux de [Biernacki et al. \(2002\)](#), deux types de contraintes sont considérées pour transférer les proportions des classes d'un modèle de mélange : soit elles sont identiques sur les modèles *Source* et *Cible* dans la famille de modèles $(M_j)_{j \in \{1, \dots, 5\}}$, soit elles sont différentes dans la famille de modèles $(pM_j)_{j \in \{1, \dots, 5\}}$ et doivent donc être ré-estimées pour le modèle *Cible*. Dans un contexte de modèle markovien pour la classification, le champ de Markov caché \mathbf{y} contrôle d'une part la structure d'interactions spatiales des classes et d'autre part les proportions de mélange. Il convient alors d'appliquer les mêmes types de contraintes sur le paramètre β contrôlant la distribution du champ de Markov. En termes de définition des modèles, cela revient à considérer les deux options suivantes :

- $\beta_T = \beta_S$: les images *Source* et *Cible* ont la même structure spatiale.
- $\beta_T \neq \beta_S$: les images *Source* et *Cible* ont des structures spatiales différentes, β_T doit être réestimée.

Le transfert d'apprentissage pour la classification dans le cadre markovien pourra se faire selon dix modèles récapitulés dans le tableau 3.2. Ce tableau fournit aussi le nombre de paramètres libres de chaque modèle.

TABLE 3.2 – Modèles de transfert dans le cadre markovien

	$M1$	$M2$	$M3$	$M4$	$M5$
Lien entre les covariables	$D_k = \mathbb{I}_d$	$D_k = \alpha \mathbb{I}_d$	$D_k = D$	$D_k = \alpha_k \mathbb{I}_d$	$D_k = D_k$
Structures spatiales	$\beta_T = \beta_S$	$\beta_T = \beta_S$	$\beta_T = \beta_S$	$\beta_T = \beta_S$	$\beta_T = \beta_S$
Nombre de paramètres	0	1	d	K	$d.K$
	$\beta M1$	$\beta M2$	$\beta M3$	$\beta M4$	$\beta M5$
Lien entre les covariables	$D_k = \mathbb{I}_d$	$D_k = \alpha \mathbb{I}_d$	$D_k = D$	$D_k = \alpha_k \mathbb{I}_d$	$D_k = D_k$
Structures spatiales	$\beta_T \neq \beta_S$	$\beta_T \neq \beta_S$	$\beta_T \neq \beta_S$	$\beta_T \neq \beta_S$	$\beta_T \neq \beta_S$
Nombre de paramètres	$\frac{K(K+1)}{2}$	$1 + \frac{K(K+1)}{2}$	$d + \frac{K(K+1)}{2}$	$K + \frac{K(K+1)}{2}$	$d.K + \frac{K(K+1)}{2}$

3.2.2 Estimation de paramètres

L'algorithme trNREM que nous introduisons dans cette section est une extension dans le cadre du transfert d'apprentissage de l'algorithme NREM proposé par [Ce-
leux et al. \(2003\)](#) et présenté en section 3.1.3.4. En utilisant une approximation en champ moyen qui permet de fixer les interactions entre des pixels voisins, il sera possible d'estimer les paramètres $\psi = (\beta, D_1, \dots, D_K)$ du couple de variables $(\mathbf{x}_T, \mathbf{y}_T)$. Cette

approximation permettra de considérer les labels de classes comme une collection de variables indépendantes $\mathbf{y}_T = \{y_{T_1}, y_{T_2}, \dots, y_{T_n}\}$ et donc de se ramener à un modèle de mélange indépendant sur lequel il est possible d'utiliser l'algorithme EM pour estimer les paramètres.

La forme générale de l'algorithme NREM n'est pas changée dans le cadre du transfert d'apprentissage. L'algorithme consiste toujours à alterner une phase (NR) de restauration des voisinages où la distribution $P(\mathbf{y}_T|\beta)$ est approximée à l'aide du champ voisin \mathbf{y}_T^{mf} , et une phase (EM) d'estimation des paramètres du modèle de transfert à l'aide de la procédure classique de [Dempster et al. \(1977\)](#). Dans le cadre de l'algorithme trNREM l'ensemble des paramètres à estimer se réduit à $\psi = (\beta, D_1, \dots, D_K)$ ou à $\psi = (D_1, \dots, D_K)$ dans le cas le plus simple. Les composantes du mélange gaussien *Cible* peuvent être en effet être exprimées en fonction des paramètres connus du modèles *Source* ($\mu_k; \Sigma_k$) et des matrices D_k à estimer :

$$f(x_i; D_k) = \frac{1}{\sqrt{(2\pi)^p |D_k \Sigma_k^S D_k|}} \exp \left(-\frac{1}{2} (x_i - D_k \mu_k^S)^t (D_k \Sigma_k^S D_k)^{-1} (x_i - D_k \mu_k^S) \right) \quad (3.50)$$

En conséquence les changements apparaissent simplement dans la phase (M).

3.2.2.1 Forme générale de l'algorithme trNREM

La forme générale de l'algorithme trNREM est présentée ici dans le cas où l'estimation des paramètres se fait pour l'un des modèles de la famille $(\beta M_j)_{j \in \{1, \dots, 5\}}$. Pour des questions de lisibilité des notations, la lettre $_T$ sera omise pour faire référence aux paramètres du modèle *Cible*, mais la lettre $_S$ sera maintenue pour faire référence aux paramètres du modèle *Source*. En utilisant les covariables x_i du modèle *Cible* et en partant de paramètres initiaux fixés à $\psi^{(0)} = (\beta_S, \{D_k\}_{k \in \{1, \dots, K\}})$ où $\forall k \in \{1, \dots, K\}, D_k = \mathbb{I}_d$, l'itération $(q+1)$ de l'algorithme consiste à réaliser les phases suivantes :

(NR) *Neighbourhood restoration* : En utilisant les paramètres courants $\psi^{(q)}$ et les observations x_i , calculer un champ voisin \mathbf{y}^{mf} à partir de la distribution $\mathbf{y}|\mathbf{x}$ approximée par

$$P_{\mathbf{y}^{mf}}(\mathbf{y}|\mathbf{x}, \psi^{(q)}) = \prod_{i \in I} P(y_i | x_i, \mathbf{y}_{N(i)}^{mf}) \quad (3.51)$$

où la fonction énergie de $\mathbf{y}|\mathbf{x}$ est donnée par

$$H^{mf}(\mathbf{y}|\mathbf{x}, \psi) = H^{mf}(\mathbf{y}|\beta) - \sum_{i \in I} \sum_{k=1}^K y_{ik} \log f(x_i; D_k) \quad (3.52)$$

(EM) *Estimation des paramètres* : En utilisant les paramètres courants $\psi^{(q)}$, estimer les nouveaux paramètres $\psi^{(q+1)}$ du modèle de distribution jointe donnée par

$$P_{\mathbf{y}^{mf}}(\mathbf{x}, \mathbf{y}|\psi) = \prod_{i \in I} \prod_{k=1}^K (\pi_{ik} f(x_i; D_k))^{y_{ik}} \quad \text{où } \pi_{ik} = P(y_{ik} = 1 | x_i, \mathbf{y}_{N(i)}^{mf}, \beta) \quad (3.53)$$

Sachant que sous une approximation en champ moyen l'espérance à maximiser est de la forme

$$\mathcal{Q}(\psi|\psi^{(q)}) = \mathbb{E}_{\psi^{(q)}}[\log P_{\mathbf{y}^{mf}}(\mathbf{x}, \mathbf{y}|\psi)|\mathbf{x}] \quad (3.54)$$

$$= \sum_{i \in I} \sum_{k=1}^K t_{ik}^{(q)} \log[\pi_{ik}^{(q)} f(x_i; D_k^{(q)})] \quad (3.55)$$

où $t_{ik}^{(q)}$ sont les probabilités *a posteriori*.

(E) Calcul des probabilités *a posteriori* pour tous $i \in I$ et $k \in \{1, \dots, K\}$:

$$t_{ik}^{(q)} = \frac{\pi_{ik}^{(q)} f(x_i; D_k^{(q)})}{\sum_{l=1}^K \pi_{il}^{(q)} f(x_i; D_l^{(q)})} \quad (3.56)$$

(M) Mise à jour des paramètres $\psi = (\beta, D_1, \dots, D_K)$ en résolvant les problèmes de maximisation suivants :

$$\beta^{(q+1)} = \arg \max_{\beta} \sum_{i \in I} \sum_{k=1}^K t_{ik}^{(q)} \log P(y_{ik} = 1 | x_i, \mathbf{y}_{N(i)}^{mf}, \beta) \quad (3.57)$$

$$\{D_k\}_{k \in \{1, \dots, K\}}^{(q+1)} = \arg \max_{\{D_k\}} \sum_{i \in I} \sum_{k=1}^K t_{ik}^{(q)} \log f(x_i; D_k) \quad (3.58)$$

Après m itérations, lorsqu'il n'y a plus de changements significatifs entre deux itérations, les labels associés à l'image I_T peuvent être restaurés en utilisant la règle du MAP. Pour chaque pixel $i \in I$ la classe la plus probable $y_i = k$, étant donnée l'observation x_i , est donnée par :

$$\forall i \in I, \quad y_i^{MAP} = \arg \max_k P_{\mathbf{y}^{mf}}(y_i = k | x_i) = \arg \max_k t_{ik} \quad (3.59)$$

3.2.2.2 Implémentation

Dans cette section, nous abordons l'implémentation des différentes étapes de trN-REM puisque différentes versions existent selon le modèle de transfert choisi. En pratique, l'étape (NR) peut être effectuée selon l'une des trois procédures en champ moyen, modal ou simulé présentées en section 3.1.3.4. Comme pour l'algorithme NREM, elle est réalisée sur le champ conditionnel $\mathbf{y}|\mathbf{x}$ afin d'obtenir une restauration plus fiable que celle qui pourrait être obtenue en prenant simplement le champ \mathbf{y} .

Lors de l'application de la procédure (EM), l'étape E reste la même pour tous les modèles $(M_j)_{j \in \{1, \dots, 5\}}$ ou $(\beta M_j)_{j \in \{1, \dots, 5\}}$. Mais l'étape M varie selon le modèle de transfert choisi puisque les paramètres à estimer varient selon les contraintes. Les paramètres de tous les modèles peuvent être appris à l'aide d'un algorithme de descente ou d'ascendance du gradient (Boyd et Vandenberghe, 2004), bien que des solutions analytiques existent pour certains modèles.

Paramètre β Pour les modèles de la famille $(M_j)_{j \in \{1, \dots, 5\}}$ il n'y a pas besoin d'estimer β puisque $\beta_T = \beta_S$. La matrice β étant fixée, l'étape M se réduit simplement à la maximisation de l'équation (3.58). Lorsque les modèles de la famille $(\beta M_j)_{j \in \{1, \dots, 5\}}$ sont considérés, la matrice symétrique β d'un modèle de Potts peut être estimée en utilisant un algorithme de descente de gradient pour maximiser la fonction suivante :

$$g(\beta) = \sum_{i \in I} \sum_{k \in \mathcal{K}} t_{ik} [\log(W^{mf}(\beta)^{-1}) + \log(\sum_{j \sim i} y_i^t \beta y_j^{mf})] \quad (3.60)$$

Modèle M1 ($D_k = \mathbb{I}_d$). L'utilisation du modèle $(\beta M1)$ implique que seule la matrice β doit être estimée comme montré précédemment. L'étape M se réduit à l'équation (3.57). Et lorsque le modèle $(M1)$ est utilisé, il n'y a pas de paramètre à estimer puisqu'ils sont tous identiques à ceux du modèle *Source*. Seule l'équation (3.56) doit être calculée afin de produire la classification.

Modèle M2 ($D_k = \alpha \mathbb{I}_d$) / **Modèle M4** ($D_k = \alpha_k \mathbb{I}_d$). Pour ces modèles, [Biernacki et al. \(2002\)](#) ont montré que maximiser l'expression (3.58) sous la contrainte du modèle M2 et respectivement du modèle M4 conduit à résoudre des équations du second degré ayant respectivement une solution non négative de la forme :

$$\alpha = \frac{1}{2} \left\{ -\frac{\sum_k n_k \bar{x}^t \Sigma_k^{S-1} \mu_k^S}{nd} + \sqrt{\left(\frac{\sum_k n_k \bar{x}^t \Sigma_k^{S-1} \mu_k^S}{nd} \right)^2 + 4 \frac{\sum_{k,i} t_{ik} x_i^t \Sigma_k^{S-1} x_i}{nd}} \right\} \quad (3.61)$$

$$\alpha_k = \frac{1}{2} \left\{ -\frac{\bar{x}^t \Sigma_k^{S-1} \mu_k^S}{d} + \sqrt{\left(\frac{\bar{x}^t \Sigma_k^{S-1} \mu_k^S}{d} \right)^2 + 4 \frac{\sum_i t_{ik} x_i^t \Sigma_k^{S-1} x_i}{n_k d}} \right\} \quad (3.62)$$

où $n_k = \sum_{i=1}^n t_{ik}$ et $\bar{x} = \sum_{i=1}^n \frac{t_{ik} x_i}{n_k}$.

Modèle M3 ($D_k = D$) / **Modèle M5** ($D_k = D_k$). Pour les modèles $(M3)$ et $(M5)$ il n'existe pas de solution analytique donnant les matrices D_k . Mais [Biernacki et al. \(2002\)](#) montrent que le problème de maximisation de l'équation (3.58) est équivalent à trouver l'unique minimum des fonctions suivantes :

$$h(D) = -2n \log |D^{-1}| + \sum_i x_i^t D^{-1} \sum_k t_{ik} \Sigma_k^{S-1} D^{-1} x_i - 2 \sum_k n_k \mu_k^{S^t} \Sigma_k^{S-1} D^{-1} \bar{x} + cst, \quad (3.63)$$

et

$$h(D_1, \dots, D_K) = \sum_k h_k(D_k) \quad (3.64)$$

où

$$h_k(D_k) = -2n_k \log |D_k^{-1}| + \sum_i x_i D_k^{-1} \left(t_{ik} \Sigma_k^{S-1} \right) D_k^{-1} x_i - 2n_k \mu_k^{S^t} \Sigma_k^{S-1} D_k^{-1} \bar{x} + cst. \quad (3.65)$$

En prenant en compte ces solutions, nous avons pu implémenter facilement les différents modèles de l'algorithme trNREM sous Matlab. en utilisant la toolbox UGM⁴ développée par Mark Schmidt pour diverses tâches reposant sur le formalisme des modèles graphiques probabilistes. Cette toolbox qui contient des implémentations de l'échantillonneur de Gibbs, de l'algorithme en champ moyen et de l'algorithme ICM fournit les codes nécessaires pour implémenter l'étape NR de l'algorithme trNREM et générer des images simulées.

La question qui reste ouverte à ce stade concerne la sélection du modèle le plus adapté à chaque situation de transfert. Nous aborderons cette question et les différentes solutions envisageables en présentant les résultats d'expérimentations de l'algorithme.

3.2.2.3 Variante de l'algorithme pour l'apprentissage supervisé et l'estimation du modèle Source

Comme nous l'avons indiqué en début de section, les tâches qui nous intéressent sont des tâches de classification supervisée à partir d'une image *Source* dont les classes sont déterminées par la connaissance d'un ensemble d'apprentissage. L'algorithme de transfert d'apprentissage présenté ici repose sur la connaissance complète de l'image *Source* : toutes les valeurs du couple $(\mathbf{x}_S, \mathbf{y}_S)$ sont supposées connues. Pour connaître les paramètres du modèle markovien associé à cette image nous proposons aussi d'utiliser une variante de l'algorithme NREM pour apprendre les paramètres $\psi_S = (\beta_S, \mu_k^S, \Sigma_k^S)$, $\forall k \in \{1, \dots, K\}$ dans un cadre supervisé. L'idée est alors d'employer une procédure d'apprentissage en deux phases à l'exemple de celle utilisée dans les travaux de [Pyun et al. \(2007\)](#), qui proposent de réaliser une classification par pixel supervisée à l'aide d'un modèle markovien de percolation de liens (*Bond percolation model*).

Nous proposons de décomposer un processus d'apprentissage supervisé en deux phases :

1. L'apprentissage des paramètres du modèle de mélange gaussien : les paramètres (μ_k^S, Σ_k^S) peuvent être estimés en utilisant les labels disponibles dans un cadre supervisé ou semi-supervisé avec un algorithme EM ([Côme, 2009](#)).
2. L'apprentissage des paramètres du champ de Markov caché : la matrice β_S peut être estimée en utilisant l'algorithme trNREM avec le modèle $(\beta M1)$. En effet, même si cette procédure a été présentée pour réaliser une tâche de transfert d'apprentissage, ce modèle peut être utilisé pour apprendre β_S puisque tous les autres paramètres du modèle sont fixes.

4. <http://www.di.ens.fr/~mschmidt/Software/UGM.html>

3.2.3 Expérimentations sur données simulées

Dans cette section nous illustrons les performances de l'algorithme trNREM pour un problème de classification par pixel appliqué à des images simulées I_S et I_T . Les images utilisées pour ces expériences sont générées en utilisant l'échantillonneur de Gibbs (Geman et Geman, 1984) avec des paramètres connus, ce qui permet de contrôler les structures des images. L'objectif est d'évaluer le comportement de l'algorithme sur des cas simples, avant de confirmer ce comportement sur des données réelles plus complexes et parfois plus difficiles à interpréter.

3.2.3.1 Simulation d'images à l'aide de l'échantillonneur de Gibbs

Les images synthétiques sont générées en utilisant l'échantillonneur de Gibbs selon un modèle de Potts pour le champ discret \mathbf{y} codant la structure des classes. Le champ est donc paramétré par une matrice symétrique β de taille $K \times K$ mais par souci de simplicité, les images simulées seront paramétrées par une matrice de la forme :

$$\beta = \begin{pmatrix} A & b & \dots & b \\ b & A & \dots & b \\ \dots & \dots & \dots & \dots \\ b & b & \dots & A \end{pmatrix} \quad \text{où } A > b, (A, b) \in \mathbb{R}^2 \quad (3.66)$$

La figure 3.2 présente deux images de taille 100×100 générées en utilisant ce modèle pour $K = 6$ classes et une structure spatiale définie sur un système de voisinage 4D. Elles correspondent à deux types de structures spatiales imposées par la matrice β :

- Image à structure proche de l'aléatoire : β_{rand} définie par $A = 1$; $b = 0.8$.
- Image à groupes concentrés : β_{struc} définie par $A = 1$; $b = 0.3$.

Les covariables $\mathbf{x} = \{X_1, \dots, X_d\}$ associées à ces images sont générées par des distributions gaussiennes conditionnelles aux classes, telles que si le pixel i est dans la classe k , le vecteur $x_i = (x_{i1}, \dots, x_{ip})$ est tiré aléatoirement de la distribution gaussienne paramétrée par $\mathcal{N}(\mu_k, \Sigma_k)$. Ici nous considérons le cas $p = 3$ covariables. Le tableau 3.3 présente les paramètres utilisés pour générer les covariables associées à l'image que nous considérerons comme l'image *Source* I_S . Les paramètres de l'image *Cible* I_T sont eux déterminés par la relation stochastique les liant à ceux de I_S via le modèle (M2) : $D_k = \alpha \text{Id}$, où $\alpha = 0.7$.

3.2.3.2 Évaluation des gains du transfert d'apprentissage

En reprenant les images générées selon les modèles précédents, nous évaluerons le gain d'utilisation de l'algorithme markovien de transfert d'apprentissage pour réaliser

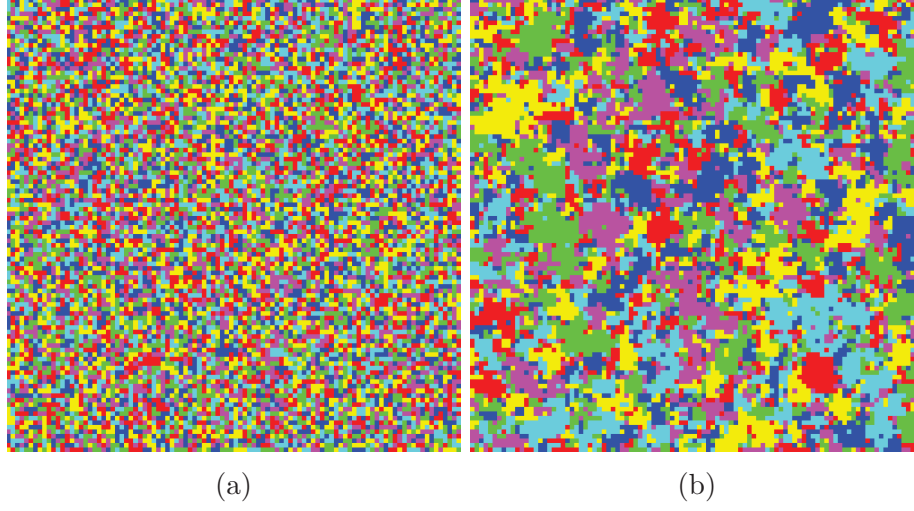


FIGURE 3.2 – Images simulées en utilisant un modèle de Potts à 6 classes : (a) Image à structure aléatoire (β_{rand}), (b) Image à groupes concentrés (β_{struc})

TABLE 3.3 – Paramètres de distribution des covariables de l'image I_S

μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
$(1.5 \ -2 \ 3)$	$(-1 \ 1 \ 2)$	$(3 \ -3 \ 1)$	$(5 \ 2 \ -1)$	$(-2 \ 0.5 \ 3.5)$	$(0.2 \ 3 \ -2)$
σ_1	σ_2	σ_3	σ_4	σ_5	σ_6
$\begin{pmatrix} 1.2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$	$\begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$	$\begin{pmatrix} 0.8 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.8 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}$

une classification par pixel d'une image *Cible* donnée. Les résultats sont présentés sur quatre expériences de transfert correspondant à deux cas où I_S et I_T ont la même structure spatiale et deux cas où elles ont des structures spatiales différentes. Rappelons que les covariables $\mathbf{x}_S|\mathbf{y}_S$ et $\mathbf{x}_T|\mathbf{y}_T$ sont simulées selon un lien de distribution correspondant au modèle $M2$.

Nous présentons les résultats de l'algorithme trNREM pour les modèles ($M1, \beta M1$) et ($M2, \beta M2$). De plus nous comparons ces résultats à ceux obtenus pour les modèles ($M1, pM2$) de l'algorithme de transfert d'apprentissage sur des mélanges gaussiens de [Biernacki et al. \(2002\)](#), qui ne prend pas en compte la dépendance des données. Cet algorithme sera noté trGMM dans la suite. L'algorithme trNREM est lancé avec la procédure en champ simulé pour l'étape NR et tourne pendant 100 itérations, tandis que trGMM tourne jusqu'à convergence.

Pour évaluer les gains apportés par ces algorithmes, nous disposons pour les images *Cible* d'une vérité terrain qui nous a permis de calculer des erreurs de classification dans des situations optimales d'apprentissage auxquelles seront confrontés tous les résultats

d'apprentissage par transfert. Les erreurs de classification de ces situations d'apprentissage sont données dans le tableau 3.4, sachant que pour obtenir des performances stables et pour s'abstraire du fait que les images sont simulées aléatoirement, chaque image a été simulée cent fois avec les mêmes paramètres. Les résultats présentés dans cette section correspondent donc systématiquement à des moyennes sur ces cent simulations.

TABLE 3.4 – Erreurs de classification par apprentissage supervisé

	GMM		HMM	
	Moyenne	<i>Écart-type</i>	Moyenne	<i>Écart-type</i>
I_T^{struc}	0,12	0,005	0,07	0,004
I_T^{rand}	0,12	0,003	0,12	0,003

Le tableau 3.5 présente les taux d'erreur de classification dans les différentes situations d'apprentissage, avec ou sans adaptation aux caractéristiques de l'image I_T . Quel que soit l'algorithme utilisé, il est clairement plus avantageux de se baser sur des techniques de transfert d'apprentissage que de se baser sur l'apprentissage supervisé de I_S appliqué sans adaptation à I_T . En regardant les différents taux d'erreur associés aux situations d'apprentissage sans transfert, il apparaît que les performances de l'algorithme trNREM sont moins bonnes que celles de l'algorithme trGMM. Le lissage spatial apporté par la matrice β du champ de Markov caché semble dégrader les performances du classifieur lorsque le modèle de mélange est incorrect. De plus, si l'on tente de ré-apprendre ce paramètre spatial, alors l'erreur de classification augmente encore. Cette situation traduit la nécessité d'appliquer un lissage spatial seulement lorsque le modèle de mélange est correctement spécifié.

Cependant, en regardant les erreurs de classification associées aux situations d'apprentissage par transfert (les modèles $pM2$, $M2$, et $\beta M2$), il apparaît cette fois que l'algorithme trNREM accomplit de meilleures performances que l'algorithme trGMM. Mais seulement lorsque le paramètre spatial β représente de façon adéquate la véritable structure spatiale de l'image *Cible*. En effet sur ces images simulées l'utilisation du bon modèle de transfert, c'est-à-dire $\beta M2$, conduit à une précision de classification aussi bonne que celle obtenue dans les conditions optimales d'apprentissage supervisé (cf. tableau 3.4). En regardant les cas présentés plus en détail, on se rend compte que le gain apporté par trNREM face à trGMM n'est pas significatif lorsque l'image *Cible* I_T possède une structure spatiale proche de l'aléatoire. Il l'est cependant lorsque I_T possède une structure spatiale marquée, appelant à la modélisation de ces dépendances spatiales.

TABLE 3.5 – Taux d'erreur de classification sur l'image *Cible* dans différentes conditions d'apprentissage

		Sans transfert			Avec transfert		
		trGMM (M1)	trNREM (M1)	trNREM ($\beta M1$)	trGMM ($pM2$)	trNREM (M2)	trNREM ($\beta M2$)
$I_S^{rand} \Rightarrow I_T^{rand}$	Moyenne	0.18	0.19	0.25	0.12	0.12	0.12
	Écart-type	0.005	0.006	0.006	0.004	0.003	0.004
$I_S^{struc} \Rightarrow I_T^{struc}$	Moyenne	0.18	0.18	0.21	0.12	0.07	0.08
	Écart-type	0.010	0.014	0.016	0.005	0.004	0.004
$I_S^{struc} \Rightarrow I_T^{rand}$	Moyenne	0.18	0.23	0.25	0.12	0.13	0.12
	Écart-type	0.007	0.006	0.005	0.004	0.004	0.004
$I_S^{rand} \Rightarrow I_T^{struc}$	Moyenne	0.18	0.18	0.22	0.12	0.10	0.08
	Écart-type	0.009	0.012	0.016	0.005	0.004	0.003

Le tableau 3.6 donne plus de précision sur ces résultats en comparant les distances entre les estimations des paramètres et les vrais paramètres utilisés pour simuler les images. Lorsque les modèles de la famille $M2$ sont utilisés, les algorithmes trGMM et trNREM fournissent des estimations équivalentes des paramètres de mélange (μ_k, Σ_k) . Ces estimations sont systématiquement plus proches des vrais paramètres que celles obtenues avec les modèles de la famille $M1$ qui correspond à l'apprentissage sans transfert de ces paramètres de mélange. On constate que les différences de performances sont liées à l'utilisation et à l'estimation du paramètre β :

- Lorsque les structures spatiales des images *Source* et *Cible* sont similaires il semble, d'après les deux premières expériences, plus approprié d'apprendre la matrice β sur l'image qui a les composantes de mélange les plus séparables (dans notre cas I_S).
- Lorsque l'image à classer possède une structure spatiale proche de l'aléatoire (1^{ère} et 3^{ème} expériences), même si β est correctement estimé, l'intégration d'une structure spatiale est insignifiante. Les classifieurs construits par les algorithmes trGMM ou trNREM sont équivalents.
- Lorsque les structures spatiales des images *Source* et *Cible* diffèrent (3^{ème} et 4^{ème} expériences), il est plus efficace de construire un classifieur en s'adaptant à la structure de l'image *Cible* : structure concentrée ou aléatoire. L'utilisation d'un paramètre β correctement appris entraîne une amélioration des performances de classification.

Ces quatre expériences soulignent le fait qu'il est plus efficace d'apprendre un classifieur pour une nouvelle image dans un cadre de transfert d'apprentissage. Elles soulignent aussi la complexité des modèles markoviens qui sont sensibles à la précision de

TABLE 3.6 – Distances aux vrais paramètres de l'image *Cible*

		Sans transfert			Avec transfert		
		trGMM	trNREM	trNREM	trGMM	trNREM	trNREM
		(M1)	(M1)	($\beta M1$)	($pM2$)	(M2)	($\beta M2$)
$I_S^{rand} \Rightarrow I_T^{rand}$	Moyenne	-	0.005	0.040	-	0.005	0.007
	β Écart-type	-	0.002	0.002	-	0.002	0.002
	μ_k Moyenne	7.16	7.16	7.16	0.84	0.84	0.84
	μ_k Écart-type	0.06	0.06	0.06	0.05	0.05	0.05
	Σ_k Moyenne	2.13	2.13	2.13	1.14	1.14	1.14
	Σ_k Écart-type	0.10	0.10	0.10	0.05	0.05	0.05
$I_S^{struc} \Rightarrow I_T^{struc}$	Moyenne	-	0.004	0.055	-	0.004	0.006
	β Écart-type	-	0.001	0.006	-	0.001	0.001
	μ_k Moyenne	7.16	7.16	7.16	0.84	0.84	0.79
	μ_k Écart-type	0.06	0.06	0.06	0.06	0.06	0.05
	Σ_k Moyenne	2.11	2.11	2.11	1.15	1.15	1.16
	Σ_k Écart-type	0.10	0.10	0.10	0.05	0.05	0.05
$I_S^{struc} \Rightarrow I_T^{rand}$	β Moyenne	-	0.044	0.041	-	0.044	0.007
	β Écart-type	-	0.001	0.003	-	0.001	0.002
	μ_k Moyenne	7.14	7.14	7.14	0.83	0.83	0.83
	μ_k Écart-type	0.06	0.06	0.06	0.05	0.05	0.05
	Σ_k Moyenne	2.13	2.13	2.13	1.14	1.14	1.14
	Σ_k Écart-type	0.10	0.10	0.10	0.05	0.05	0.05
$I_S^{rand} \Rightarrow I_T^{struc}$	β Moyenne	-	0.043	0.055	-	0.043	0.006
	β Écart-type	-	0.001	0.005	-	0.001	0.001
	μ_k Moyenne	7.15	7.15	7.15	0.83	0.83	0.79
	μ_k Écart-type	0.06	0.06	0.06	0.05	0.05	0.05
	Σ_k Moyenne	2.12	2.12	2.12	1.14	1.14	1.16
	Σ_k Écart-type	0.10	0.10	0.10	0.05	0.05	0.06

l'ensemble des paramètres : si β est appris avec un mauvais modèle de mélange, l'algorithme produira un paramètre inexact entraînant rapidement le classifieur dans une mauvaise direction. Mais si tous les paramètres sont appris et transférés correctement, alors il devient possible d'obtenir des résultats de classification aussi bons que ceux qui pourraient être obtenus dans une situation optimale d'apprentissage supervisé.

3.3 Application : la construction d'une base de tissus urbains

Les premiers travaux sur grilles radioconcentriques (cf. chapitre 2) nous ont permis d'identifier une nomenclature de tissus urbains qui fait sens pour décrire des espaces urbains caractérisés par différentes oppositions structurantes d'un point de vue morphologique. Dans le but de pouvoir traiter un plus grand nombre d'espaces urbains et

selon des hypothèses plus relâchées d'organisations spatiales (notamment l'existence de multiples centres au sein d'un même espace urbain), nous souhaitons maintenant transposer sur les grilles régulières l'*a priori* morphologique qui était contenu dans le format des grilles radioconcentriques. Nous proposons de l'introduire sur ces grilles sous la forme d'un *a priori* probabiliste contenu dans la structure des dépendances spatiales et dans la nomenclature de classes recherchée. Pour cela nous utiliserons l'algorithme trNREM pour constituer une base de données des tissus urbains en tirant avantage du transfert d'apprentissage.

Lorsqu'une seule aire urbaine est étudiée, la classification des tissus urbains peut être aisément réalisée par classification supervisée. Dans ce cas, des labels sont collectés sur certaines parties de l'aire urbaine afin de déterminer les prototypes statistiques des tissus urbains selon des variables d'études pertinentes pour construire un modèle prédictif. Mais lorsque plusieurs aires urbaines sont étudiées, le coût d'acquisition de labels sur chaque espace à classer devient un obstacle qui peut être surmonté par l'emploi de méthodes de transfert. L'algorithme trNREM trouve alors tout son intérêt pour la constitution d'une base de données homogénéisée des tissus urbains, en permettant de s'adapter au contexte local de chaque aire urbaine qui peut entraîner une dérive de la distribution statistique des variables d'études (figure 3.3). Le transfert d'apprentissage permettra d'améliorer les résultats des classifications en supposant par exemple que les prototypes de classes de tissus urbains, qui sont déterminés spatialement et en fonction

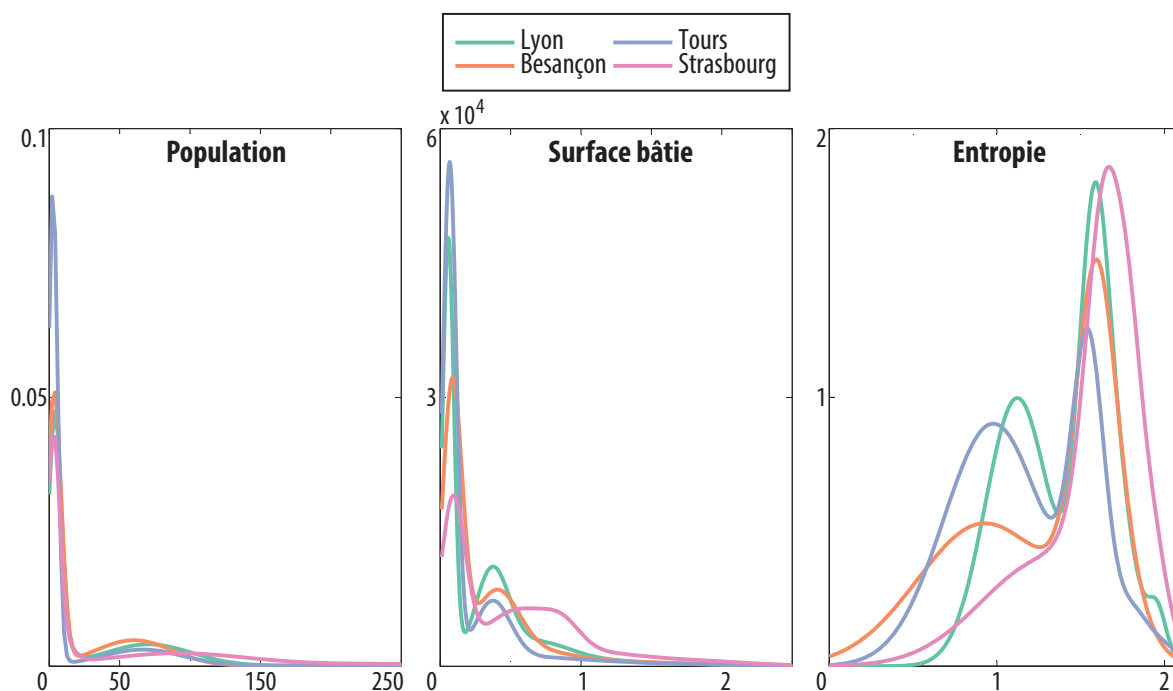


FIGURE 3.3 – Densités de probabilité marginales des variables de population, surface bâtie et entropie pour des modèles de mélange appris sur les différentes sources d'apprentissage qui seront utilisées pour constituer la base de tissus urbains

de variables d'études telles que la population et les surfaces bâties, varieront selon la taille des agglomérations, selon les spécificités locales d'aménagement et contraintes topographiques, ou plus généralement selon l'histoire urbaine de chaque territoire.

3.3.1 Mise en œuvre de la base de données

La base de données des tissus urbains est constituée sur l'ensemble des 50 plus grandes aires urbaines françaises, exception faite de Paris, récapitulées dans le tableau 3.21. Le gigantisme de l'aire urbaine parisienne, comparée au reste du système d'aires urbaines françaises nous pousse à l'écarter de cette étude. Que ce soit pour l'application de la méthode de transfert d'apprentissage ou pour respecter les objectifs plus généraux de formalisation d'une typologie des configurations urbaines françaises, la spécificité de la configuration parisienne tend à la traiter comme un cas à part sans chercher à apprendre sa structure par les moyens développés pour les autres aires urbaines de l'échantillon.

Nous détaillons ici les éléments nécessaires à l'apprentissage par transfert, à savoir la composition des domaines *Sources* et *Cibles*, puis nous définirons les paramètres de mise en œuvre de l'algorithme utilisé pour constituer la base de tissus urbains.

3.3.1.1 Les bases d'apprentissages

Pour traiter cet ensemble de 50 aires urbaines, il est nécessaire de choisir les aires urbaines qui serviront de source d'apprentissage. Au-delà de la nécessité d'adaptation aux spécificités d'organisation propres à chaque aire urbaine, certaines tendances plus globales et communes à un plus grand nombre d'aires urbaines pourraient avoir un impact sur la qualité du transfert d'apprentissage. On peut par exemple supposer que la flexibilité des modèles de transfert choisis pourrait trouver ses limites si les aires urbaines *Source* et *Cible* ont des tailles trop différentes. Ou encore que la flexibilité des modèles sera restreinte si les aires urbaines ont des formes d'habitat beaucoup trop différentes en termes de concentration et dispersion des bâtiments, liées notamment aux particularités régionales se retrouvant dans les formes d'habitats.

Plusieurs *Sources* d'apprentissage sont définies et testées pour construire la base de tissus urbains. Quatre *Sources* nous permettront de tester la flexibilité des modèles et de déterminer quels sont les facteurs influant le plus la qualité du transfert :

- Lyon : modèle de la grande ville, aux formes d'habitats peu concentrées
- Strasbourg : modèle de la grande ville, aux formes d'habitats concentrées
- Tours : modèle de la ville moyenne, aux formes d'habitats peu concentrées
- Besançon : modèle de la petite ville, aux formes d'habitats concentrées

A partir des grilles régulières de ces quatre aires urbaines, des travaux sur grilles radioconcentriques pour Tours et Besançon, ou de connaissances externes sur la trame bâtie (BD Topo, cartes IGN, etc.) pour Lyon et Strasbourg, quatre bases de labels composées d’un ensemble de labels d’entraînement et d’un ensemble de labels de test sont constituées. L’ensemble d’entraînement permettra de construire un classifieur markovien selon la procédure d’apprentissage supervisé présentée en section 3.2.2.3, tandis que l’ensemble de test permettra d’évaluer les différents classifieurs construits par transfert d’apprentissage. Les caractéristiques de ces bases sont données dans le tableau 3.7.

TABLE 3.7 – Bases de labels (nombre de pixels)

		JEUX D’ENTRAÎNEMENT				JEUX DE TEST			
		Lyon	Strasbourg	Tours	Besançon	Lyon	Strasbourg	Tours	Besançon
1	Habitat dispersé	950	360	1500	1600	50	25	30	25
2	Habitat individuel	630	270	500	1000	50	25	30	25
3	Habitat Mixte	150	80	45	80	50	25	30	25
4	Grands Ensembles	30	30	20	20	50	25	30	25
5	Centre Historique	45	40	25	20	50	25	30	25
6	Zone d’activité	200	150	150	300	50	25	30	25
Total		2005	930	2240	3020	300	150	180	150

3.3.1.2 Sélection des variables d’intérêt

Dans les travaux préparatoires sur les grilles radioconcentriques nous avons fait le choix de travailler sur trois variables, la population, les surfaces bâties et les surfaces moyennes, suffisamment discriminantes pour déterminer la composition en tissus urbains des trois agglomérations étudiées. Dans le cadre de l’extension de ce travail de classification au traitement d’une plus grande variété d’aires urbaines, nous proposons d’étendre le choix des variables d’études, sélectionnées parmi l’ensemble des variables qu’il nous est possible de calculer à partir des données de la BD Topo concernant les bâtiments (cf. 2.1.2).

En utilisant les bases d’apprentissages dont nous disposons maintenant, il est en effet possible d’appliquer une procédure de sélection de variables. De telles techniques permettent de déterminer les variables les plus importantes pour une tâche donnée de classification ou de régression en cherchant à éliminer les variables non pertinentes ou redondantes (Guyon et Elisseeff, 2003). Nous nous intéressons ici à une approche de type ordonnancement de variables qui consiste à attribuer à chaque variable j , ($j = 1, \dots, p$) un score calculé à partir d’un ensemble d’apprentissage $\mathcal{A} = (x_i, y_i), i = 1, \dots, n$ où x_i est un vecteur de p variables prédictives et y_i est la variable de labels à prédire.

Dans ce cadre, l’algorithme des forêts aléatoires de Breiman (2001) offre la possibilité de réaliser des applications de sélection de variables par ordonnancement (Diaz-Uriarte et de Andres, 2006; Genuer et al., 2010). Le calcul de tels scores repose sur l’utilisation des erreurs de classification associées aux différents arbres composant la

forêt. La méthode de *bagging* employée pour construire chaque arbre t permet de déterminer un ensemble OOB_t (le *Out Of Bag Sample*) correspondant aux données qui n'ont pas été utilisées pour construire le prédicteur λ_t . Cet ensemble qui contient environ un tiers des observations du jeu d'apprentissage (Breiman, 2001), permet de calculer une erreur de classification empirique donnée par :

$$errOOB_t = \sum_{y_k \in OOB_t} \mathbb{1}_{y_k \neq \hat{y}_k} \quad (3.67)$$

En se basant sur cette erreur, il est possible de calculer un score, appelé l'importance des variables VI , pour chaque variable x_j de l'échantillon. Pour chaque échantillon OOB_t les valeurs de la i -ème variable sont permutées aléatoirement de façon à obtenir un échantillon perturbé \widetilde{OOB}_t^j . L'importance d'une variable correspond alors à l'augmentation moyenne de l'erreur $errOOB_t$ d'un arbre t après permutation, et est donnée pour une forêt par :

$$VI(X_i) = \frac{1}{M} \sum_t (err\widetilde{OOB}_t^j - errOOB_t) \quad (3.68)$$

L'objectif est d'utiliser le score VI pour déterminer un ordonnancement des variables et sélectionner celles qui seront les plus pertinentes pour mener des tâches de classification des tissus urbains par transfert d'apprentissage. Pour cela, nous nous servons des labels disponibles pour construire quatre jeux de données équilibrés par classes

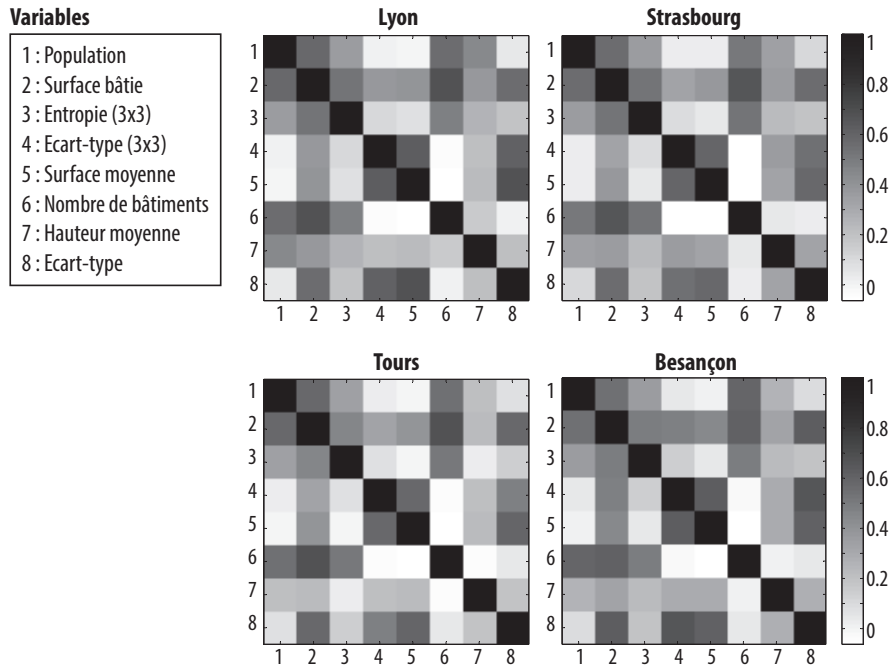


FIGURE 3.4 – Valeurs absolues des corrélations entre variables pour les quatre départements incluant les sources d'apprentissage

sur les grilles de Lyon, de Strasbourg, de Tours et de Besançon. L'importance des $p = 8$ variables suivantes est testée :

1. la population par maille (POP)
2. la surface cumulée des bâtiments par maille (SURF)
3. l'entropie calculée dans un voisinage de 3x3 mailles (ENT3x3)
4. l'écart-type de la surface des bâtiments dans un voisinage de 3x3 mailles (STDEV3x3)
5. la surface moyenne des bâtiments par maille (S_MOY)
6. le nombre de bâtiments par maille (NB_BAT)
7. la hauteur moyenne des bâtiments par maille (H_MOY)
8. l'écart-type de la surface des bâtiments par maille (STDEV)

La figure 3.4 présente les corrélations entre ces variables pour les quatre départements correspondant aux agglomérations servant de sources d'apprentissage. On retrouve des corrélations plus ou moins fortes, similaires sur les quatre sources, sachant que les variables présentées au chapitre 2 qui étaient trop fortement corrélées aux autres (notamment le volume) ont été ignorées ici. Les variables conservées présentent toujours des liaisons significatives entre elles. Ces liaisons partielles indiquent que ces variables peuvent toutefois apporter une information différente et complémentaire, qui sera peut-être utile pour la classification des tissus urbains.

Pour déterminer un ordonnancement global de ces variables, nous avons appliqué la procédure suivante : une forêt de 5000 arbres est calculée sur chacun des quatre jeux de données, en considérant des sous-ensembles de \sqrt{p} variables à chaque nœud de décision. Les scores VI qui sont obtenus pour chaque jeu d'apprentissage sont ensuite moyennés pour obtenir un score VI global donné en figure 3.5. Considérer les différentes *Source* d'apprentissage séparément avant de les agréger pour obtenir un score global tient

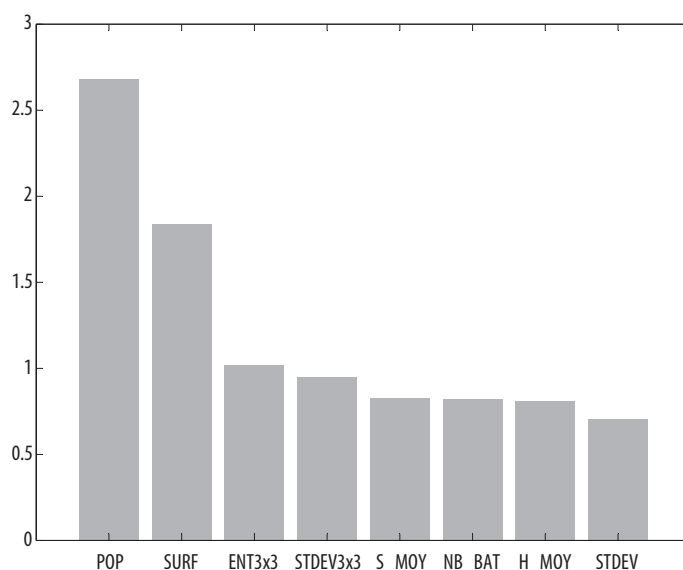


FIGURE 3.5 – Importances globales des variables

d'une volonté de ne pas privilégier une des *Sources* de transfert, qui rappelons-le ont été délibérément choisies de façon à être représentatives de différentes structururations urbaines et statistiques. L'hypothèse est que cet ordonnancement global nous permettra d'identifier les variables qui seront les plus pertinentes pour traiter la majorité des agglomérations par transfert d'apprentissage.

Pour obtenir une idée plus précise du sous-ensemble de variables à conserver nous donnons en figure 3.6 les erreurs OOB de classification pour les quatre jeux de données, en fonction du nombre de variables ordonnées par importance globale. Cette figure confirme ce que l'ordonnancement global indiquait déjà, à savoir que la population (POP) et les surfaces bâties cumulées (SURF) jouent un rôle nettement plus important pour réaliser les classifications des tissus urbains. Ces deux variables permettent une très bonne identification des tissus urbains, puisque les taux d'erreur OOB sont à 5% ou moins sur les quatre jeux de données. Les variables suivantes sont plus difficiles à discriminer tant en termes d'importances globales que d'erreur OOB. Il est intéressant de noter que les indicateurs de diversité ou de complexité des formes bâties apparaissent directement à la suite des variables de densité. Le pouvoir discriminant de ces variables est plus faible, mais elles apportent toutefois une certaine stabilité en termes d'erreur de classification qui continue à décroître, mais beaucoup moins rapidement. Ces variables, même si elles présentent une utilité moindre pour la prédiction, peuvent toutefois présenter une utilité en termes d'interprétation des classes des tissus urbains.

Dans l'esprit de conserver une description des tissus urbains en termes de densité

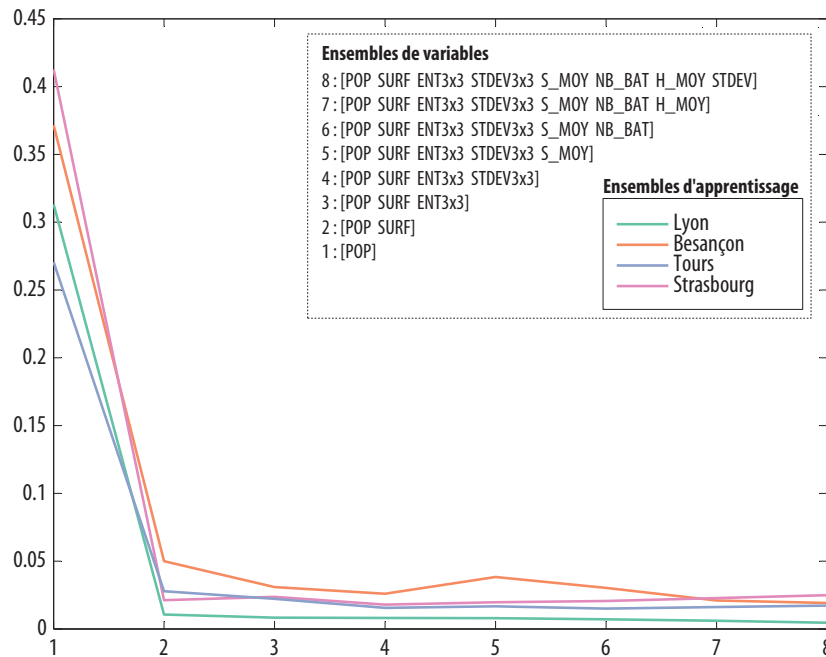


FIGURE 3.6 – Erreurs OOB de classifications sur les quatre jeux d'apprentissage

TABLE 3.8 – Valeurs moyennes de classes pour les quatre jeux d'apprentissages

	Lyon			Strasbourg		
	POP	SURF	ENT3x3	POP	SURF	ENT3x3
Habitat individuel dispersé	4	575	1,12	3	916	1,30
Habitat individuel groupé	71	3679	1,58	90	4927	1,65
Habitat mixte et dense	329	7395	1,72	253	8737	1,81
Grands ensembles	960	5857	1,72	830	6886	1,79
Centre historique	954	20774	1,96	527	19067	1,95
Zone d'activité	5	8987	1,66	3	10238	1,67

	Tours			Besançon		
	POP	SURF	ENT3x3	POP	SURF	ENT3x3
Habitat individuel dispersé	3	673	0,98	4	869	0,93
Habitat individuel groupé	66	3706	1,54	60	4040	1,60
Habitat mixte et dense	244	7868	1,83	239	8398	1,78
Grands ensembles	745	6325	1,79	709	6960	1,80
Centre historique	342	17921	2,01	459	20080	1,88
Zone d'activité	2	9397	1,62	2	9530	1,59

et de diversité des formes bâties, nous faisons le choix de conserver les trois premières variables : la population et les surfaces bâties cumulée, indicateurs de densité, et l'entropie des surfaces bâties, indicateur de diversité. Ces trois variables permettent de constituer des prototypes de classes pour nos quatre jeux d'apprentissage qui sont donnés en tableau 3.8, tableau à mettre en perspective avec le tableau 2.2 des hypothèses qualitatives formulées dans la phase préliminaire de travail sur les grilles radioconcentriques.

3.3.1.3 Paramètres d'application de l'algorithme

Pour appliquer l'algorithme trNREM, il est nécessaire de fixer plusieurs paramètres en entrée qui auront un impact sur les sorties.

Pixels à classer Les premières expérimentations nous ont conduit à choisir d'appliquer l'algorithme non pas sur les images entières des aires urbaines, mais sur un sous-ensemble de pixels $I0$ tel que :

$$I0 = \{i \in I | \text{SURF}(i) > 0 \wedge \text{INFRA}(i) < \text{SURF}(i)\} \quad (3.69)$$

où SURF est l'image des surfaces bâties par mailles de 200 mètres et INFRA est l'image des surfaces d'infrastructures de transport et de cours d'eau par maille de 200 mètres (cf. chapitre 2).

Pour faciliter le traitement des données, nous choisissons donc de ne pas inclure les pixels qui ne sont pas intéressants pour la classification des tissus urbains : nous

TABLE 3.9 – Erreurs de classification selon les voisinages (4-D ou 8-D contraint) et l'estimation du champ voisin

SOURCE	MODÈLE	CIBLE	8-D			4-D		
			Champ simulé*	Champ modal	Champ moyen	Champ simulé*	Champ modal	Champ moyen
Tours	$\beta M3$	Lyon	0,01	0,01	0,01	0,01	0,01	0,01
Strasbourg	$\beta M4$		0,03	0,02	0,02	0,03	0,03	0,02
Besançon	$M3$		0,04	0,04	0,04	0,04	0,04	0,04
Tours	$\beta M4$	Strasbourg	0,03	0,03	0,03	0,03	0,03	0,03
Besançon	$M2$		0,06	0,04	0,04	0,06	0,05	0,05
Lyon	$\beta M2$		0,09	0,09	0,09	0,09	0,09	0,08
Strasbourg	$\beta M4$	Tours	0,04	0,03	0,03	0,03	0,03	0,02
Besançon	$M2$		0,01	0,02	0,02	0,01	0,02	0,02
Lyon	$\beta M3$		0,06	0,08	0,07	0,07	0,08	0,07
Tours	$\beta M3$	Besançon	0,04	0,03	0,04	0,03	0,03	0,03
Strasbourg	$M2$		0,09	0,08	0,09	0,08	0,09	0,07
Lyon	$\beta M3$		0,07	0,07	0,07	0,07	0,07	0,06

* Les résultats de la procédure en champ simulé correspondent à des moyennes sur 20 lancers de l'algorithme trNREM

ignorons tous les pixels correspondant à des surfaces non bâties, et tous ceux correspondant à des pixels majoritairement occupés par des infrastructures de transport ou cours d'eau. Ce filtrage aura l'avantage d'exclure un grand nombre de pixels qui en plus d'être non intéressants, peuvent ralentir l'algorithme.

Système de voisinage Le fait de travailler sur un sous-ensemble de l'image amène à revoir le système de voisinages employé pour modéliser les interactions spatiales entre classes. Le système de voisinage qui pouvait correspondre aux 4 ou 8 plus proches voisins est maintenant réduit : selon le choix initial, 4-D ou 8-D, le système de voisinage est contraint aux plus proches voisins restant dans l'ensemble I_0 . Certains pixels peuvent se retrouver sans voisins, ce qui conduit à ne pas prendre en compte d'interactions spatiales pour estimer la classe de ces pixels lors du décodage du champ conditionnel $\mathbf{x}|\mathbf{y}$. Pour cette application, la structure spatiale imposée par la nomenclature des tissus urbains entraîne que ces pixels correspondent majoritairement à des pixels du type habitat individuel dispersé. Le tableau 3.9 montre qu'en termes d'erreurs de classification il n'y a pas de grandes différences entre les modèles appris par transfert avec un système de voisinage 4-D ou 8-D. Puisque le système de voisinage 4-D semble apporter un léger gain et qu'il implique un temps de calcul plus court, nous utiliserons ce système de voisinage pour constituer la base de données.

Champ voisin Pour lancer l'algorithme trNREM il est nécessaire de choisir une des trois procédures pour estimer le champ voisin. Les expériences de transfert entre les différentes images *Source* indiquent que les trois procédures fournissent des résultats

à peu près équivalents en termes d'erreur de classification. Toutefois, l'algorithme en champ simulé semble moins adapté pour cette application puisque cette procédure stochastique donne des résultats moins stables. Bien que l'initialisation soit systématiquement la même (les paramètres de l'image *Source*), l'algorithme est entraîné dans des optima locaux variables et parfois de moins bonne qualité que les optima identifiés par la procédure en champ modal ou en champ moyen. Les résultats donnés dans le tableau 3.9 indiquent que la procédure en champ simulé est en moyenne légèrement moins bonne que les deux autres procédures. La procédure en champ moyen qui semble apporter un léger gain par rapport à la procédure en champ modal sera utilisée pour constituer la base de données.

3.3.2 Évaluation de la méthode

Nous nous proposons d'évaluer ici les performances de l'algorithme trNREM selon les différentes sources d'apprentissage, en reprenant les hypothèses selon lesquelles les différences de taille et de forme de bâti peuvent impacter la qualité du transfert entre aires urbaines. Comme pour les expériences réalisées sur les images synthétiques, les performances de l'algorithme de transfert markovien trNREM seront comparées à l'algorithme de transfert gaussien trGMM. Les images présentées dans la suite correspondent aux classifications des différents types de tissus urbains introduits en section 2.2, dont les codes couleurs sont donnés dans le tableau 3.10.

3.3.2.1 Transfert entre aires urbaines de taille moyenne

La situation considérée ici correspond à l'apprentissage d'un classifieur sur une aire urbaine de taille moyenne, qui sera Besançon, à partir d'un modèle connu sur une autre aire urbaine de taille moyenne, qui sera Tours. La figure 3.7 présente les différentes images classées de l'aire urbaine de Besançon (zoomées sur l'agglomération principale) correspondant à l'apprentissage supervisé sur Besançon, à l'apprentissage sans adaptation depuis Tours, et à l'apprentissage par transfert depuis Tours selon le modèle ($\beta M3$) qui donne la meilleure erreur de classification.

TABLE 3.10 – Légende des images classées en tissus urbains








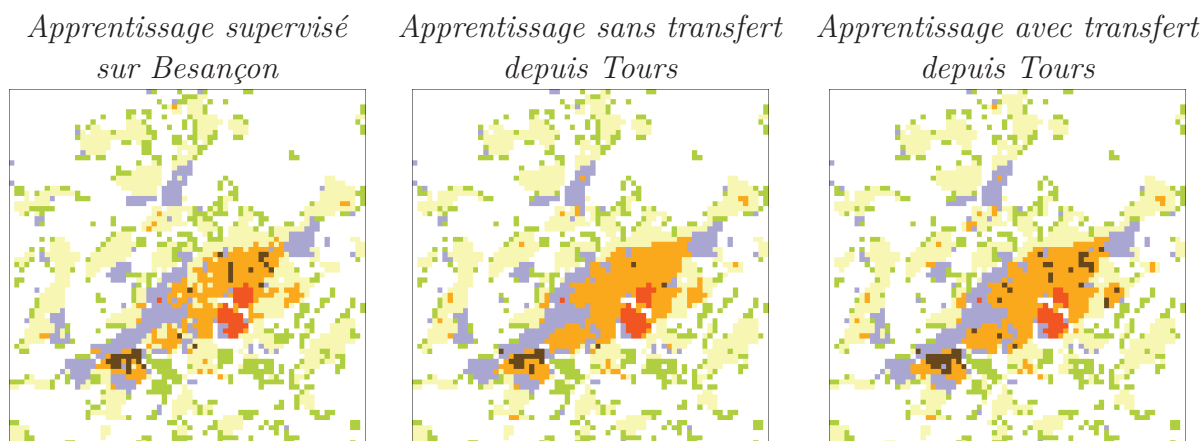
	C1. Habitat individuel dispersé
	C2. Habitat individuel groupé
	C3. Habitat mixte et dense
	C4. Habitat collectif
	C5. Centre historique
	C6. Zone d'activité
	Surfaces non bâties (non considérées lors du processus de classification)

FIGURE 3.7 – Images classées de Besançon selon différentes situations d'apprentissage depuis Tours



Le tableau 3.11 présente les erreurs de classification pour différentes situations d'apprentissage depuis le modèle *Source* construit à partir de Tours. Avec ou sans transfert, le modèle construit sur Tours apparaît assez adapté pour prédire la classification sur Besançon. Les algorithmes trNREM et trGMM appliqués sans transfert, c'est-à-dire selon la contrainte $M1$, sont capables de produire une classification fiable puisque les distorsions des covariables sont faibles sur les différentes classes. Les algorithmes de mélange gaussien GMM et trGMM sont toujours moins précis que les algorithmes markoviens, ce qui confirme l'intérêt de prendre en compte les dépendances spatiales sur ces données réelles.

La simple adaptation à la distribution des covariables sur l'aire urbaine de Besançon avec le modèle $M3$ permet déjà d'améliorer la classification. L'adaptation supplémentaire de la structure spatiale, c'est-à-dire le réapprentissage du paramètre β_T , avec le modèle $\beta M3$ permet d'améliorer encore l'apprentissage. Le tableau 3.12, présentant les matrices de confusion pour les apprentissages par trNREM, permet de préciser les gains apportés par le transfert. Le besoin d'adaptation provient de la classe C4, c'est-à-dire l'habitat collectif, qui est mal spécifiée avec le modèle $M1$ qui la sous-estime. Les modèles $M3$ et $\beta M3$ qui corrigent toutes les classes de la même façon, permettent toutefois de corriger ce problème en améliorant nettement la détection de cette classe

TABLE 3.11 – Erreurs de classification dans différentes situations d'apprentissage de Besançon depuis Tours

<i>Apprentissage supervisé sur Besançon</i>		<i>Apprentissage sans transfert depuis Tours</i>			<i>Apprentissage avec transfert depuis Tours</i>		
GMM	HMM	trGMM ($M1$)	trNREM ($M1$)	trNREM ($\beta M1$)	trGMM ($pM3$)	trNREM ($M3$)	trNREM ($\beta M3$)
0.05	0.03	0.04	0.07	0.07	0.06	0.05	0.03

TABLE 3.12 – Matrices de confusion (en nombre de pixels) des modèles de Besançon appris par trNREM depuis Tours

$(M1)$							$(M3)$							$(\beta M3)$						
	C1	C2	C3	C4	C5	C6		C1	C2	C3	C4	C5	C6		C1	C2	C3	C4	C5	C6
C1	24	1	0	0	0	0	C1	24	1	0	0	0	0	C1	24	1	0	0	0	0
C2	0	25	0	0	0	0	C2	0	25	0	0	0	0	C2	0	25	0	0	0	0
C3	0	0	25	0	0	0	C3	0	0	25	0	0	0	C3	0	0	24	1	0	0
C4	0	0	7	18	0	0	C4	0	0	3	22	0	0	C4	0	0	0	25	0	0
C5	0	0	2	0	23	0	C5	0	0	3	0	22	0	C5	0	0	2	0	23	0
C6	0	1	0	0	0	24	C6	1	0	0	0	0	24	C6	0	0	0	0	0	25

sans entraver la détection des autres classes, à l'exception peut-être de la classe C3 de l'habitat mixte.

Cette première expérience révèle l'intérêt de la prise en compte des dépendances spatiales lors du processus de transfert d'apprentissage. Elle démontre que pour des villes de taille similaire, même si les formes de bâti diffèrent légèrement, le transfert peut se faire aisément, par exemple à l'aide de la famille de modèle $(M3)$.

3.3.2.2 Transfert entre aires urbaines de grande taille

La situation de transfert présentée ici correspond au transfert entre deux aires urbaines de grande taille, illustré par l'apprentissage du classifieur de Lyon à partir de celui de Strasbourg. Les résultats de classification de l'apprentissage supervisé, de l'apprentissage sans et avec transfert par le modèle $\beta M4$ sont donnés en figure 3.8.

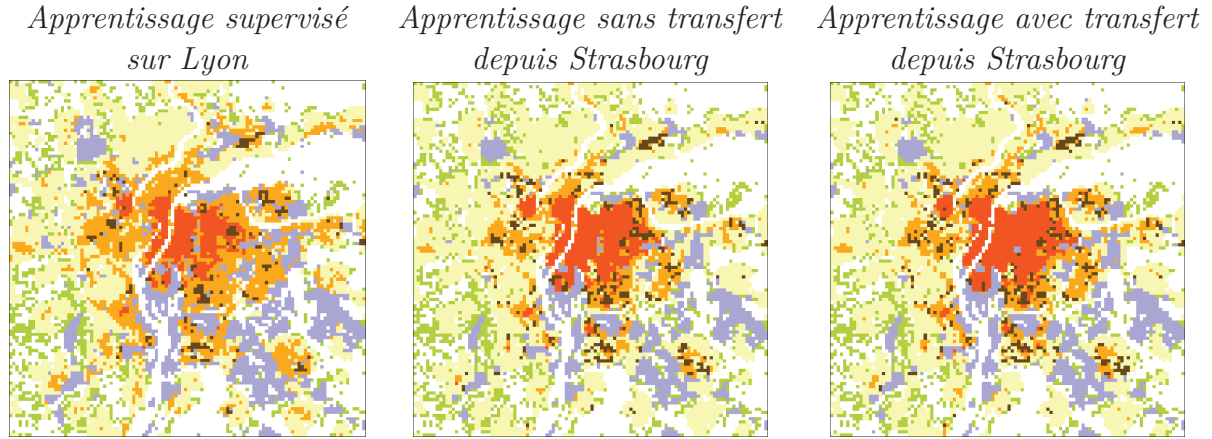
Pour le transfert entre ces aires urbaines d'assez grande taille, on se retrouve dans une situation assez similaire à la précédente : ici aussi, avec ou sans transfert, le modèle construit sur Strasbourg semble assez adapté pour produire la classification des tissus urbains de Lyon. Comme le montrent les erreurs de classification données dans le tableau 3.13 les différentes versions de l'algorithme trNREM fournissent toutes de bons résultats. Et comme dans le cas précédent l'algorithme trGMM sans dépendances spatiales fournit lui des performances bien moins bonnes.

Les différences les plus visibles entre l'apprentissage optimal et l'apprentissage avec transfert se retrouvent sur les classes d'habitat collectif et de centre historique. Les

TABLE 3.13 – Erreurs de classification dans différentes situations d'apprentissage de Lyon depuis Strasbourg

<i>Apprentissage supervisé sur Lyon</i>		<i>Apprentissage sans transfert depuis Strasbourg</i>			<i>Apprentissage avec transfert depuis Strasbourg</i>		
GMM	HMM	trGMM ($M1$)	trNREM ($M1$)	trNREM ($\beta M1$)	trGMM ($pM4$)	trNREM ($M4$)	trNREM ($\beta M4$)
0.01	0.01	0.04	0.02	0.03	0.07	0.02	0.02

FIGURE 3.8 – Images classées de Lyon selon différentes situations d'apprentissage depuis Strasbourg



matrices de confusion données dans le tableau 3.14 montrent qu'en utilisant le modèle de Strasbourg avec ou sans adaptation, la classe C2 d'habitat mixte et dense est légèrement sous-estimée au profit de la classe C4 d'habitat collectif. Bien que les modèles $M4$ et $\beta M4$ ne corrigent pas ce problème, ils améliorent la détection des autres classes, ce qui permet de conserver des modèles globalement bons.

TABLE 3.14 – Matrices de confusion (en nombre de pixels) des modèles de Lyon appris par trNREM depuis Strasbourg

$(M1)$							$(M4)$							$(\beta M4)$						
	C1	C2	C3	C4	C5	C6		C1	C2	C3	C4	C5	C6		C1	C2	C3	C4	C5	C6
C1	50	0	0	0	0	0	C1	49	1	0	0	0	0	C1	50	0	0	0	0	0
C2	0	50	0	0	0	0	C2	0	50	0	0	0	0	C2	0	50	0	0	0	0
C3	0	1	48	1	0	0	C3	0	1	47	2	0	0	C3	0	1	47	2	0	0
C4	0	0	0	49	1	0	C4	0	0	0	49	1	0	C4	0	0	0	50	0	0
C5	0	0	0	0	50	0	C5	0	0	0	0	50	0	C5	0	0	0	0	50	0
C6	3	0	0	0	0	47	C6	2	0	0	0	0	48	C6	2	1	0	0	0	47

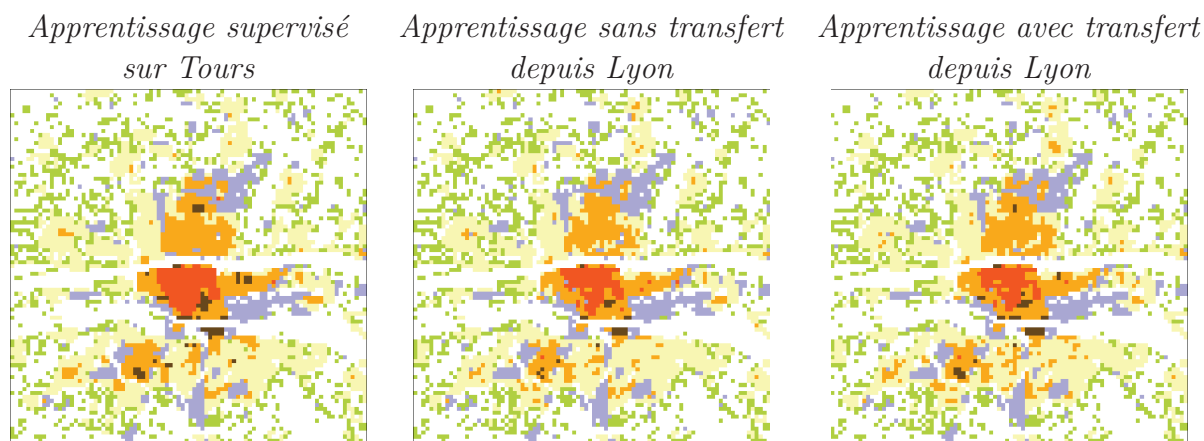
Cette deuxième expérience tend à renforcer l'idée que considérer des aires urbaines de tailles assez similaires permet de faciliter le transfert d'apprentissage. Si la différence des formes de bâti semble avoir un rôle moins important, on peut toutefois noter que dans les deux cas les meilleures erreurs de classification ont été obtenues à partir des modèles où la matrice β a été ré-apprise lors du transfert.

3.3.2.3 Transfert d'une aire urbaine de grande taille à une aire urbaine de taille moyenne

La situation de transfert présentée ici correspond à l'apprentissage d'un classifieur sur une aire urbaine de taille moyenne, Tours, à partir d'une aire urbaine de grande taille, Lyon. Les résultats de classification de l'apprentissage supervisé, de l'apprentis-

sage sans transfert et de l'apprentissage avec transfert par le modèle $\beta M3$ sont donnés en figure 3.9.

FIGURE 3.9 – Images classées de Tours selon différentes situations d'apprentissage depuis Lyon



Quelle que soit la situation d'apprentissage considérée, les erreurs de classification du tableau 3.15 indiquent qu'il est plus difficile de transférer un modèle construit sur une aire urbaine de grande taille vers une aire urbaine de taille moyenne, que de transférer un modèle sur des aires urbaines de taille similaire. Ici les distorsions de covariables sont plus grandes et les erreurs de classification lors des apprentissages sans transfert, avec ou sans prise en compte des interactions spatiales, semblent indiquer qu'un modèle construit sur une grande aire urbaine ne peut pas être appliqué directement. L'utilisation des modèles de transfert de la famille ($M3$) avec l'algorithme trNREM apportent toutefois une amélioration, ce qui n'est pas le cas avec l'algorithme trGMM. L'adaptation aux distributions des covariables permet d'améliorer la classification seulement si des interactions spatiales sont prises en compte dans le cas présent. Toutefois les erreurs de classification étant égales entre les modèles $M1$ et $\beta M1$, et entre les modèles $M3$ et $\beta M3$, on peut supposer que la matrice des interactions spatiales de Lyon est adaptée pour modéliser les interactions spatiales sur Tours. Cette situation est peut-être due aux formes de bâti globalement similaires en termes de concentration.

TABLE 3.15 – Erreurs de classification dans différentes situations d'apprentissage de Tours depuis Lyon

<i>Apprentissage supervisé sur Tours</i>		<i>Apprentissage sans transfert depuis Lyon</i>			<i>Apprentissage avec transfert depuis Lyon</i>		
GMM	HMM	trGMM ($M1$)	trNREM ($M1$)	trNREM ($\beta M1$)	trGMM ($pM3$)	trNREM ($M3$)	trNREM ($\beta M3$)
0.02	0.02	0.13	0.11	0.11	0.15	0.07	0.07

Les matrices de confusion données dans le tableau 3.16 révèlent qu'ici aussi le besoin d'adaptation vient de la classe C4 des grands ensembles. Les modèles $M3$ et $\beta M3$ permettent de corriger l'apprentissage sur cette classe tout en conservant la qualité de l'apprentissage sur les autres classes. Ces modèles qui adaptent les distributions des covariables de façon identique sur toutes les classes entraînent toutefois une dégradation de la qualité de l'apprentissage sur la classe C2 de l'habitat individuel groupé. La famille des modèles ($M5$) qui adapte le classifieur selon chaque classe et chaque variable aurait pu trouver son intérêt ici, mais dans la pratique le trop grand degré de liberté de ces modèles les entraîne dans des solutions dégradant de façon considérable les performances de classification.

TABLE 3.16 – Matrices de confusion (en nombre de pixels) des modèles de Tours appris par trNREM depuis Lyon

$(M1)$							$(M3)$							$(\beta M3)$						
	C1	C2	C3	C4	C5	C6		C1	C2	C3	C4	C5	C6		C1	C2	C3	C4	C5	C6
C1	28	2	0	0	0	0	C1	29	1	0	0	0	0	C1	29	1	0	0	0	0
C2	0	27	3	0	0	0	C2	0	25	5	0	0	0	C2	0	24	6	0	0	0
C3	0	1	29	0	0	0	C3	0	0	30	0	0	0	C3	0	0	30	0	0	0
C4	0	0	11	18	0	0	C4	0	0	3	26	0	0	C4	0	0	3	26	0	0
C5	0	0	3	0	27	0	C5	0	0	3	0	27	0	C5	0	0	3	0	27	0
C6	0	0	0	0	0	30	C6	0	0	0	0	0	30	C6	0	0	0	0	0	30

Si on compare cette expérience à la précédente, il semble que si la taille des aires urbaines joue sur la qualité du transfert, elle joue de façon moins importante que la différence des formes de bâti.

3.3.2.4 Transfert d'une aire urbaine de taille moyenne à une aire urbaine de grande taille

La situation de transfert présentée ici correspond à l'inverse de la situation précédente, à savoir l'apprentissage d'un classifieur sur une aire urbaine de grande taille, Strasbourg, à partir d'une aire urbaine de taille moyenne, Besançon. La figure 3.10 présente les classifications obtenues par apprentissage supervisé, par apprentissage sans transfert, et par apprentissage avec transfert depuis Besançon par le modèle $M4$.

Ces images et les erreurs de classification présentées dans le tableau 3.17 révèlent une plus grande facilité à réaliser le transfert entre ces aires urbaines de tailles différentes que dans le cas précédent. L'adaptation à la distribution statistique des covariables de Strasbourg n'apporte qu'un petit gain en termes d'erreur, ce qui se traduit visuellement par une surestimation de l'habitat mixte et dense. Il est intéressant de noter que comme dans le cas précédent, lorsque les aires urbaines ont des formes de bâti relativement similaires, les modèles de transfert avec ou sans ré-apprentissage de la structure spatiale de l'image *Cible* donnent des résultats similaires.

FIGURE 3.10 – Images classées de Besançon selon différentes situations d'apprentissage depuis Strasbourg

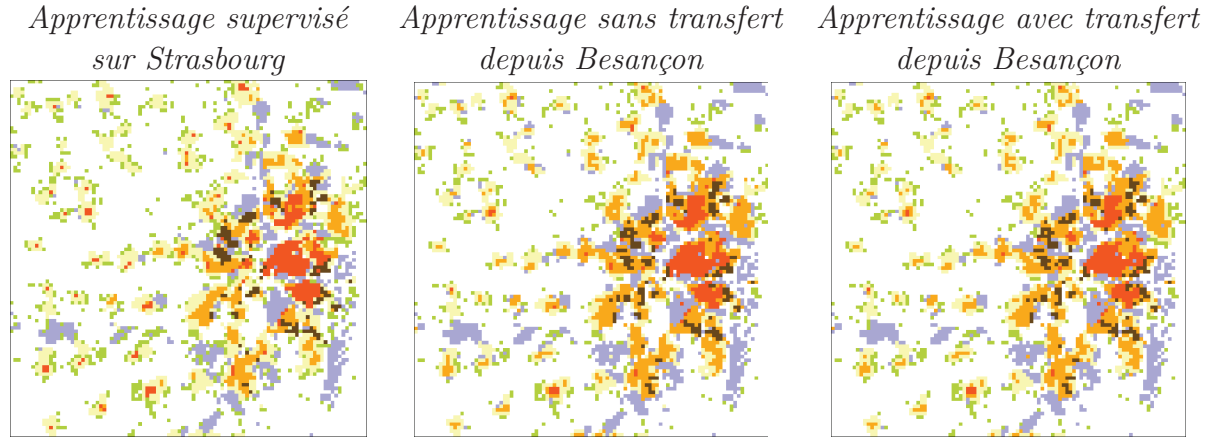


TABLE 3.17 – Matrices de confusion (en nombre de pixels) des modèles de Besançon appris par trNREM depuis Strasbourg

<i>Apprentissage supervisé sur Strasbourg</i>		<i>Apprentissage sans transfert depuis Besançon</i>			<i>Apprentissage avec transfert depuis Besançon</i>		
GMM	HMM	trGMM ($M1$)	trNREM ($M1$)	trNREM ($\beta M1$)	trGMM ($pM2$)	trNREM ($M4$)	trNREM ($\beta M4$)
0.03	0.01	0.08	0.06	0.06	0.09	0.05	0.05

Les modèles de la famille ($M4$) arrivent à produire des améliorations, mais les matrices de confusion données dans le tableau 3.18 révèlent qu'elles ne sont que partielles. Initialement le problème d'ajustement concerne les classes C2 de l'habitat individuel groupé et C3 de l'habitat mixte et dense. Les modèles $M4$ et $\beta M4$ permettent d'améliorer la situation pour la classe C2 de l'habitat individuel groupé, mais la classe C3 d'habitat mixte et dense reste surestimée de la même manière.

TABLE 3.18 – Matrices de confusion (en nombre de pixels) des modèles de Besançon appris par trNREM depuis Strasbourg

$(M1)$							$(M4)$							$(\beta M4)$						
	C1	C2	C3	C4	C5	C6		C1	C2	C3	C4	C5	C6		C1	C2	C3	C4	C5	C6
C1	22	3	0	0	0	0	C1	25	0	0	0	0	0	C1	25	0	0	0	0	0
C2	0	22	3	0	0	0	C2	0	22	3	0	0	0	C2	0	21	4	0	0	0
C3	0	0	24	1	0	0	C3	0	0	24	1	0	0	C3	0	0	24	1	0	0
C4	0	0	0	25	0	0	C4	0	0	0	25	0	0	C4	0	0	0	25	0	0
C5	0	0	0	1	24	0	C5	0	0	0	1	24	0	C5	0	0	0	1	24	0
C6	0	0	1	0	0	24	C6	0	0	2	0	0	23	C6	0	0	1	0	0	24

Ces expériences de transfert entre les sources d'apprentissage labellisées révèlent l'utilité de modèles de transfert et de la prise en compte de dépendances spatiales pour produire des classifications des tissus urbains de qualité sur n'importe quelle aire ur-

baine française. Il ressort que ces quatre sources aux caractéristiques volontairement différentes en termes de taille et de formes de bâti plus ou moins concentrées ou dispersées influencent la qualité du transfert d'apprentissage. Ces facteurs de taille et de forme ont un impact à prendre en compte pour constituer la base de tissus urbains sur des aires urbaines variées en choisissant la source la plus adaptée. De plus, les expériences menées sur ces quatre sources ne permettent pas de distinguer une famille de modèles qui soit systématiquement plus adaptée que les autres. Les distorsions de covariables ne peuvent pas être modélisées de façon identique entre toutes les aires urbaines. Il aurait alors pu être utile de considérer le modèle général $M5$ qui adapte les distributions séparément pour chaque covariable et chaque classe. Mais dans la pratique ce modèle fournit des performances de mauvaise qualité de façon systématique pour cette application. Le plus grand degré de liberté de ce modèle entraîne les classifieurs vers des solutions qui ne correspondent pas à l'apprentissage voulu des tissus urbains. Il aurait tendance à se comporter comme un classifieur construit de façon non supervisée.

3.3.2.5 Le problème de la sélection du modèle

Dans les résultats présentés sur les images *Source* sur lesquelles nous disposons de bases de labels de test, il était possible d'identifier le modèle le plus adapté en se référant à des erreurs de classification. Les différentes expériences ont montré qu'à ce stade aucun modèle n'est particulièrement plus performant pour modéliser les formes de liens entre distributions qui varient selon la paire d'aire urbaine considérée. Cette situation entraînera donc des difficultés pour la construction du reste de la base de données puisqu'il ne sera pas possible d'appliquer une procédure de sélection de modèle reposant sur l'évaluation d'erreurs de classification.

Une possibilité serait alors d'appliquer une procédure de sélection de modèle à l'aide de critères tels que BIC ou ICL (Biernacki et al., 2000). Dans le cadre d'un modèle markovien, des difficultés sont à prévoir puisque le calcul de tels critères implique le calcul de la log-vraisemblance exacte. Or, seule une version approximée de la log-vraisemblance est obtenue à cause de l'utilisation de l'approximation en champ moyen dans l'étape (NR) de l'algorithme trNREM. Pour surmonter ces difficultés Forbes et Peyrard (2003) ont introduit des critères de sélection de modèle basés sur l'approximation en champ moyen. Mais des expériences sur les images *Sources* montrent que ces versions approximées de BIC (figure 3.11) et ICL (figure 3.12) présentent des comportements instables. A ce stade nous serons donc obligé d'évaluer les résultats produits par les différents modèles en les confrontant à d'autres sources de données permettant de valider les classifications.

3.3.3 Évaluation de la base de données des tissus urbains

L'utilisation de l'algorithme trNREM avec les différents modèles de transfert nous a permis de construire une base de données composée des classifications en tissus urbains sur les 50 aires urbaines récapitulées dans le tableau 3.21 et dont la cartographie complète est donnée dans l'annexe cartographique (cf. volume II)

3.3.3.1 Comparaison à des BD MOS

L'évaluation de la qualité de l'ensemble des classifications réalisées par transfert d'apprentissage est rendue difficile par l'absence de données de validation pour l'ensemble des aires urbaines. Il n'existe pas de travaux aux objectifs similaires formalisés sur la France entière et sur des grilles régulières. La source de validation la plus proche dont nous disposons partiellement est issue des MOS (cf. encadré 2.2) réalisés pour le compte de communautés d'agglomérations, de départements ou de régions. Ces bases de données qui correspondent à des inventaires cartographiques vectoriels de tous les types d'occupation des sols sont constituées à partir de sources différentes et selon des nomenclatures différentes. Ainsi même parmi l'ensemble des MOS disponibles en open data, tous ne sont pas utilisables pour valider les classifications en tissus urbains. Par

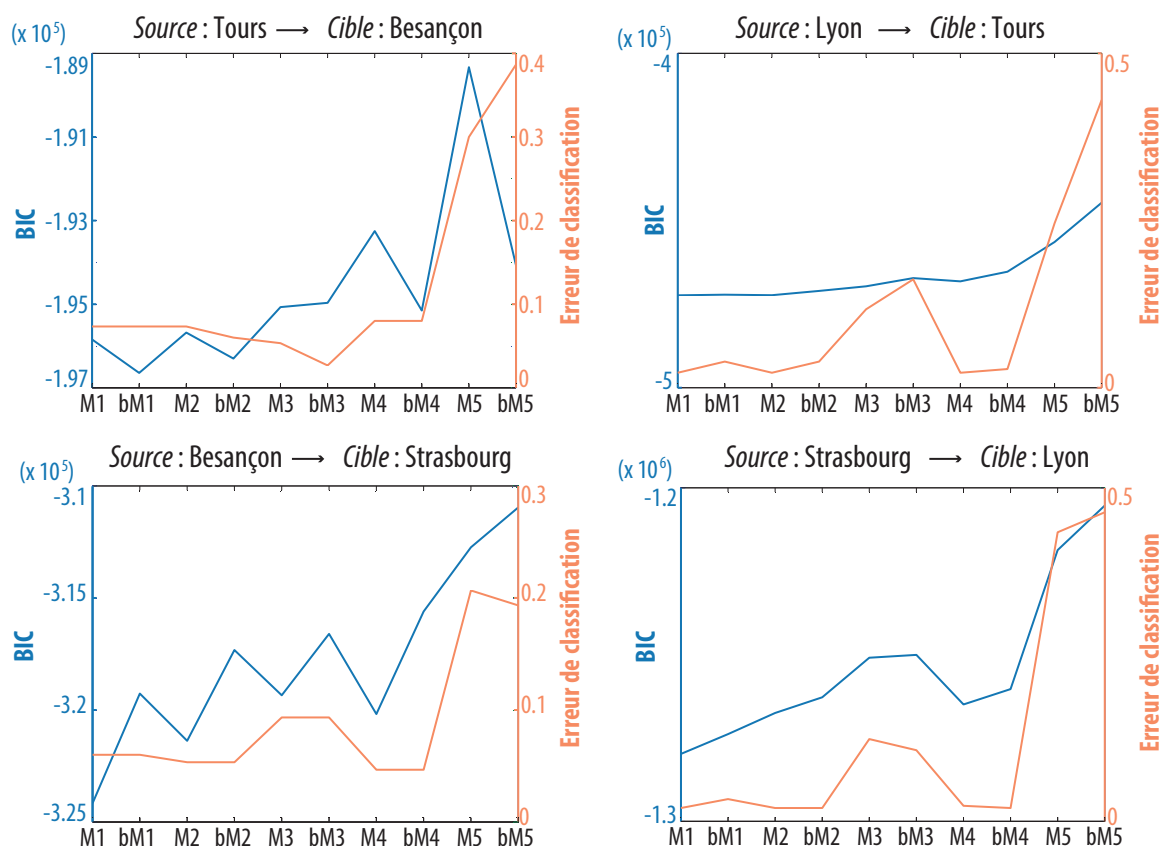


FIGURE 3.11 – Comportement de BIC face à l'erreur de classification

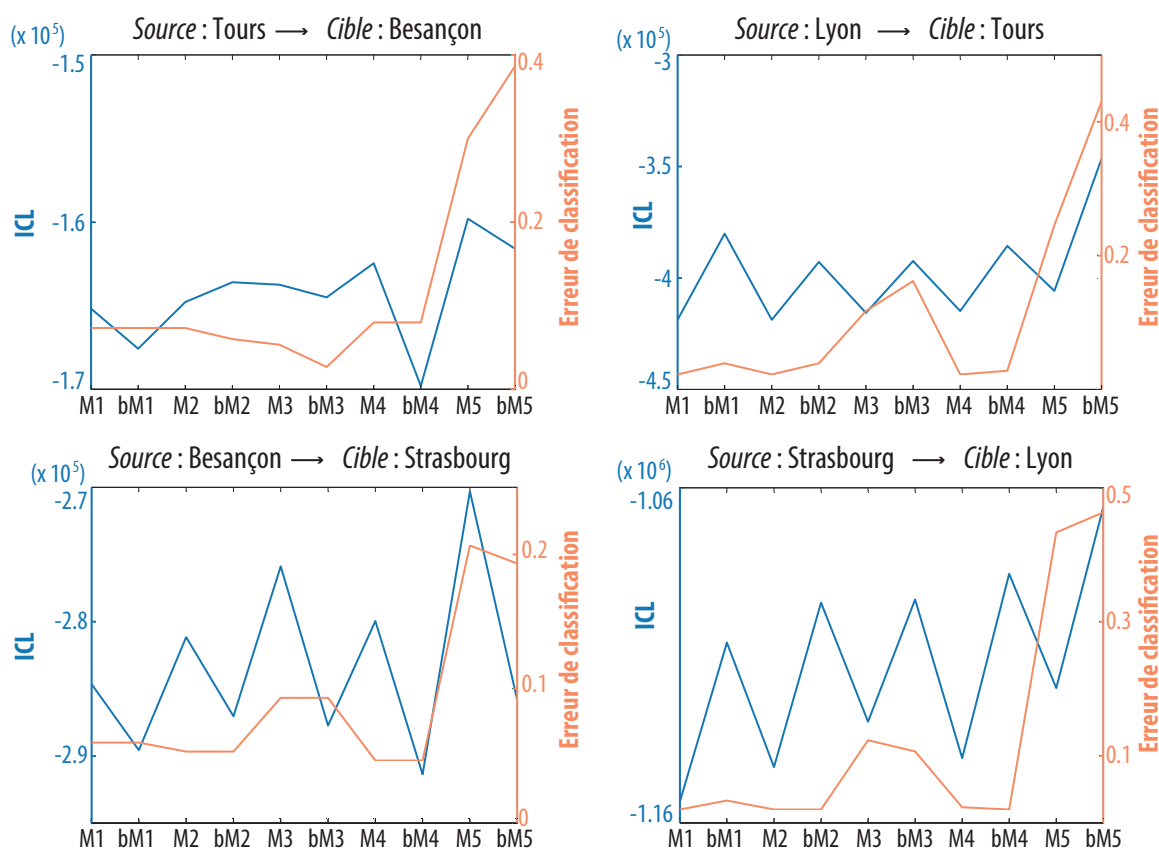


FIGURE 3.12 – Comportement de ICL face à l'erreur de classification

exemple, les MOS reposant sur la nomenclature des zones urbanisées de Corine Land Cover ne peuvent pas être exploités puisque la nomenclature, trop agrégée en milieu urbain, ne distingue que les tissus urbains continus des tissus urbains discontinus.

Certains MOS possèdent toutefois des nomenclatures qui semblent plus comparables en termes de définitions des types d'habitat. Dans ce cadre, nous prenons l'exemple de la BD MOS Loire Atlantique⁵ et de la BdOCS CIGAL de la région Alsace⁶ pour évaluer les résultats de classification obtenus avec l'algorithme trNREM sur les aires urbaines de Nantes et Strasbourg. Pour réaliser cette comparaison, nous avons d'abord filtré les BD pour ne sélectionner que les polygones associés à des types d'occupation des sols correspondant à des espaces bâtis, qu'ils soient habités ou non. Puis nous avons décomposé ces polygones sur la grille régulière à maille de 200 mètres de côté utilisée pour produire la classification par transfert. Le processus de transfert des MOS sur une grille régulière consiste simplement à attribuer le mode d'occupation des sols majoritaire en termes de surface dans une maille. Les figures 3.13 et 3.14 fournissent les résultats cartographiques des MOS agrégés et comparés aux classifications en tissus urbains de Nantes et Strasbourg.

5. <http://data.loire-atlantique.fr>

6. www.cigalsace.org

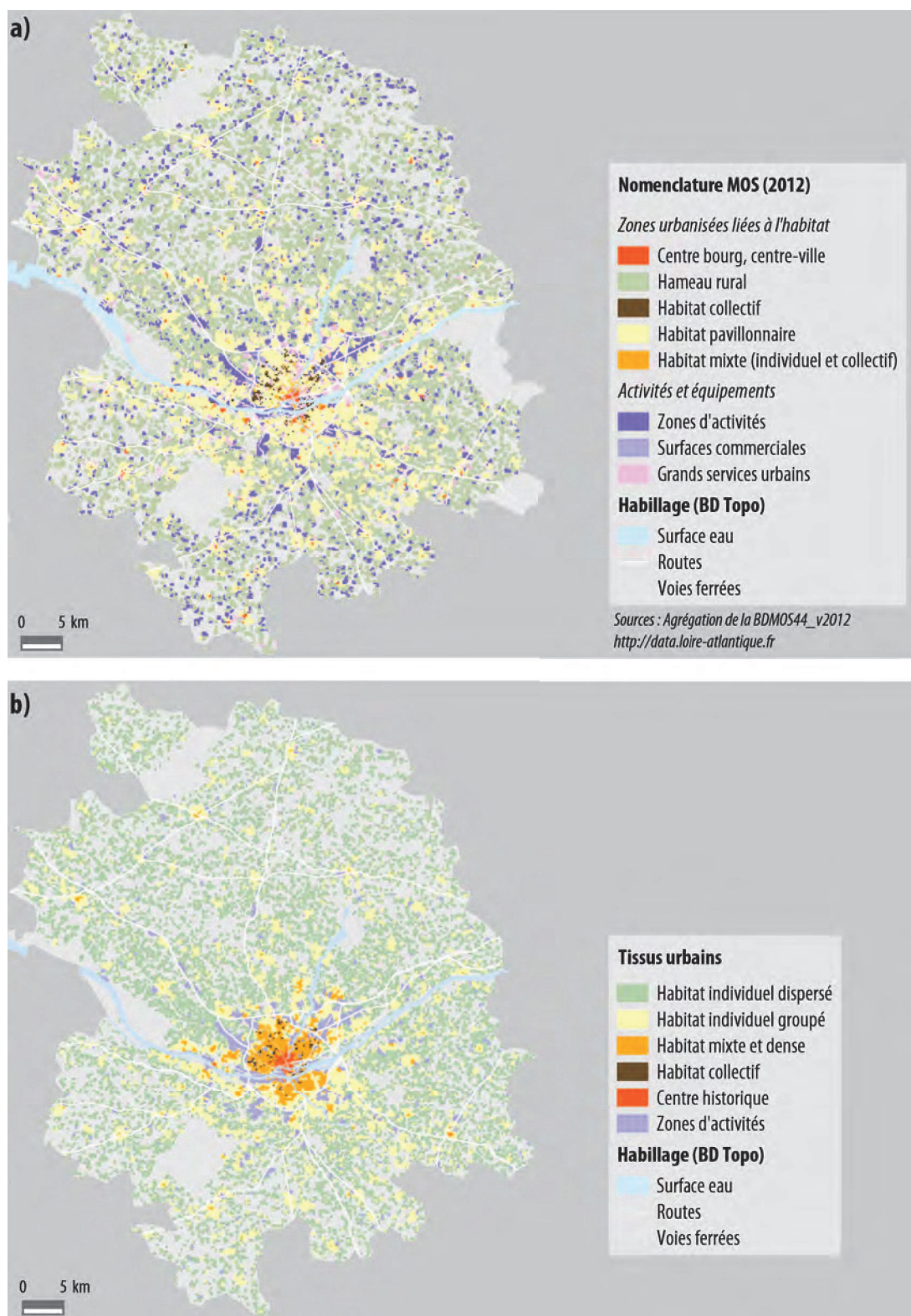


FIGURE 3.13 – Cartographie de l'aire urbaine de Nantes : a) MOS, b) classification en tissus urbains

TABLE 3.19 – Table de contingence entre les classes de tissus urbains et la BD MOS sur l'aire urbaine de Nantes

	Tissus urbains						Total
	Habitat ind. dispersé	Habitat ind. groupé	Habitat mixte et dense	Habitat collectif	Centre historique	Zone d'activité	
MOS							
Hameau rural	11204	657	0	0	0	36	11897
Habitat pavillonnaire	1842	5702	995	12	23	38	8612
Habitat mixte	2	26	87	5	26	4	150
Habitat collectif	2	9	114	82	3	11	221
Centre bourg, Centre-ville	9	97	83	0	54	0	243
Activités et équipements*	1634	414	188	7	38	1402	3683
Total	14134	7426	1467	106	144	1529	24806

* Agrégation des zones d'activités, surfaces commerciales et grands services urbains

Une comparaison quantitative des classifications en tissus urbains et des BD MOS est réalisée sous la forme de tableaux de contingences. Pour plus de lisibilité et par soucis de comparaison, les zones bâties mais non habitées (activités et équipements pour Nantes ; grandes emprises et espaces urbains spécialisés pour Strasbourg) ont été agrégées dans ces tableaux.

Le tableau de contingence 3.19 de Nantes croise ainsi les six catégories de tissus urbains à six types d'occupation des sols. La nomenclature du MOS et des tissus urbains semble très comparable, pourtant le croisement de chaque type révèle une concordance seulement partielle. L'« habitat individuel dispersé » est assez proche du type « hameau rural », de même que la classe « habitat individuel groupé » est assez proche du type « habitat pavillonnaire groupé ». Les classes d'« habitat collectif » semblent peu concordantes, mais ceci vient d'un problème (qui sera discuté dans la suite) de sous-détection de l'habitat collectif avec l'algorithme trNREM identifié pour Nantes et d'autres aires urbaines. Les autres types aux étiquettes pourtant assez semblables ne le sont pas vraiment dans la pratique pour cause de différences de définitions entre les types du MOS et les classes de tissus urbains. L'« habitat mixte » du MOS se distingue principalement de notre classe de tissu urbain d'« habitat mixte et dense » puisque la première est beaucoup plus restreinte spatialement et n'intègre pas les zones d'habitat à haute densité du centre de Nantes composées de maisons mitoyennes à deux ou trois étages et de petits immeubles. Le type « centre bourg et centre-ville » du MOS contient en fait deux catégories qui se distinguent par leurs niveaux de densités : le centre-ville ou centre historique de Nantes qui est d'ailleurs assez concordant sur les deux bases de données, et les centres de bourg qui correspondent dans notre classification à des variations locales de densité intégrées à la classe d'« habitat mixte et dense ». Enfin les « zones d'activités » sont peu concordantes avec les différents types d'activités et équipements principalement à cause de l'intégration d'exploitations agricoles aux zones d'activités et d'équipements du MOS, qui se retrouvent intégrées à l'« habitat individuel dispersé » dans la classification en tissus urbains.

Le tableau de contingence 3.20 de Strasbourg croise, lui, les six catégories de tissus

urbains aux cinq catégories d'occupation du sol du MOS, ce qui révèle qu'ici aussi les deux classifications ne sont que partiellement comparables. La catégorie d'habitat individuel du MOS correspond en fait à nos deux catégories d'habitat individuel groupé et dispersé. Comme pour Nantes, la catégorie d'habitat mixte du MOS ne correspond que très partiellement à notre catégorie d'habitat mixte et dense qui inclut sur Strasbourg des petits immeubles mêlés à de l'habitat individuel ou des lotissements d'habitat individuel à haute densité. La classe d'habitat collectif détectée via les données de densité et l'algorithme trNREM correspond mieux à la classe d'habitat collectif du MOS que pour Nantes. Mais ici aussi la concordance n'est pas parfaite puisque le prototype de classes utilisé dans l'algorithme d'apprentissage correspond à des zones à très haute densité de population, ce qui exclut certains quartiers dont la densité n'est pas très élevée sur les données. La concordance entre la classe de tissu urbain centre historique et la classe du MOS habitat continu est meilleure : les centres de bourg sont inclus dans cette classe dans le MOS, mais ils ne le sont que partiellement dans la classe centre historique. A la différence de Nantes, la source d'apprentissage que représente Strasbourg a permis d'inclure certains de ces centres de bourg dans la classe de tissu urbain centre historique puisque la très haute densité de bâtiments les rend semblables au centre historique de Strasbourg. Enfin la catégorie de tissu urbain des zones d'activités se distingue de celle du MOS qui inclut les exploitations agricoles, alors que ces dernières sont intégrées aux zones d'habitat individuel dispersé dans notre classification en tissus urbains.

TABLE 3.20 – Table de contingence entre les classes de tissus urbains et la BD MOS sur l'aire urbaine de Strasbourg

	Tissus urbains						Total
	Habitat ind. dispersé	Habitat ind. groupé	Habitat mixte et dense	Habitat collectif	Centre historique	Zone d'activité	
Habitat individuel	2095	4433	382	3	103	25	7041
Habitat mixte	7	40	35	1	39	3	125
Habitat collectif	11	80	120	112	19	4	346
Habitat continu	12	185	37	0	249	3	486
Espaces urbains spécialisés*, Grandes emprises **	891	399	71	1	36	821	2219
Total	3016	5137	645	117	446	856	10217

* Agrégation des emprises scolaires et universitaires, hospitalières, culturelles et patrimoine

** Agrégation des emprises industrielles, militaires, commerciales et artisanales, et zones d'activités tertiaires

Ces deux exemples de comparaisons entre MOS et classifications en tissus urbains révèlent la particularité de notre nomenclature en tissus urbains qui est définie en fonction de différents niveaux de densité de population et de surface bâtie, alors que les MOS sont exclusivement définis à partir d'informations relatives aux surfaces bâties. La concordance seulement partielle des MOS et des classifications en tissus urbains n'est donc pas à prendre comme un indicateur de mauvaise adaptation lors du processus d'apprentissage par transfert. Les différences repérées correspondent plutôt à des différences d'objectifs typologiques : nos classifications en tissus urbains sont consti-

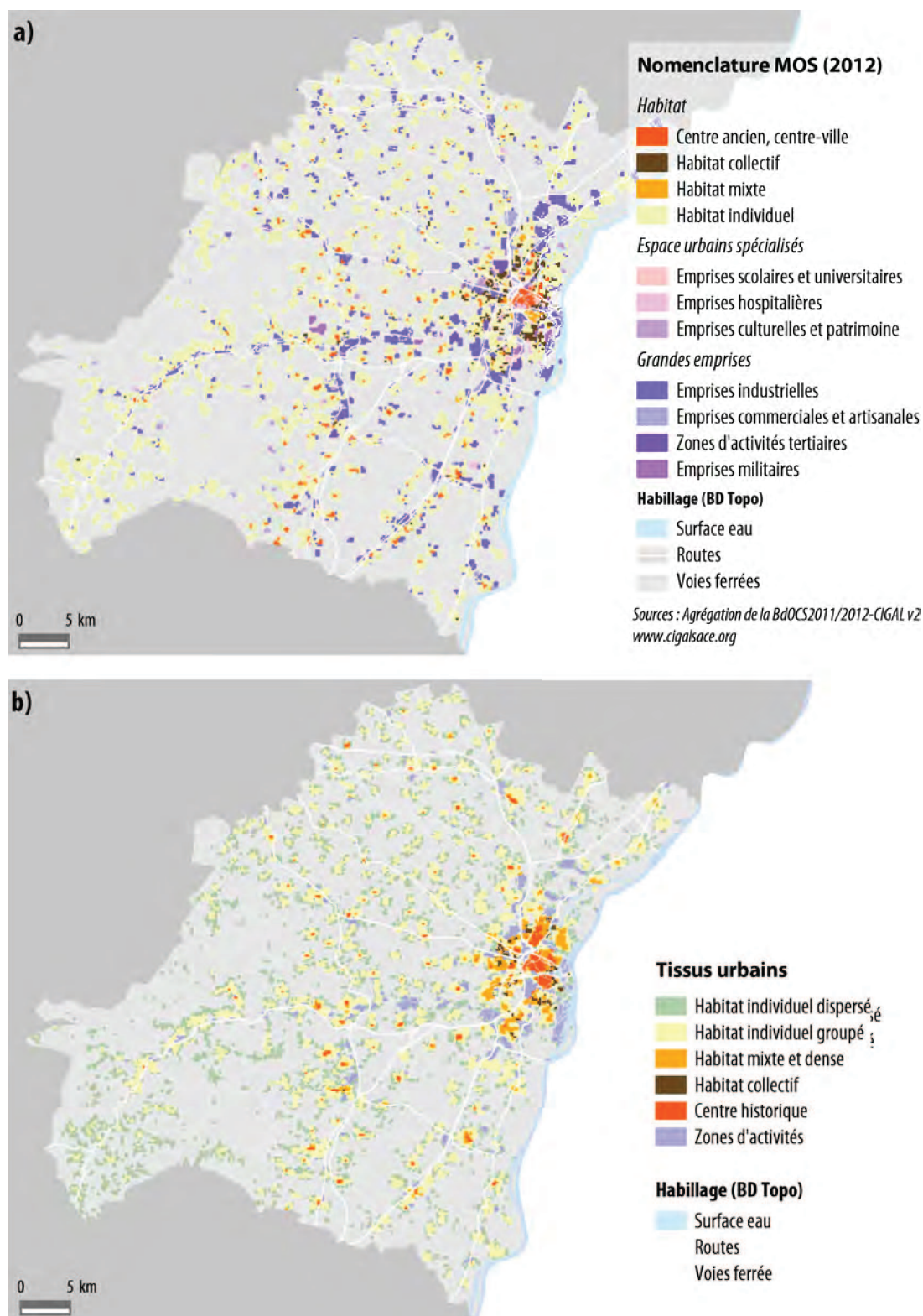


FIGURE 3.14 – Cartographie de l'aire urbaine de Strasbourg : a) MOS, b) classification en tissus urbains

tuées selon une nomenclature générale adaptée à la représentation de n'importe quel territoire urbain français, tandis que les nomenclatures variables de chaque MOS sont plus adaptées pour représenter les spécificités de chaque territoire.

3.3.3.2 Le rôle des sources et des modèles d'apprentissage

Ces deux premiers exemples de comparaisons des classifications en tissus urbains aux données externes des MOS rappellent que la constitution de cette base de données à l'échelle du territoire français repose sur la capacité à prendre en compte les spécificités locales de chaque aire urbaine lors du processus d'apprentissage pour produire des classifications adaptées. Ces spécificités ont été prises en compte en choisissant parmi les quatre sources d'apprentissage et parmi les huit modèles de transfert (la famille (M5) n'a pas été testée compte tenu des résultats expérimentaux sur les sources). Le résultat de classification qui semblait le plus pertinent a été identifié par comparaison visuelle à une superposition de données issues de la BD Topo, à la population des aires urbaines, aux MOS lorsqu'ils sont disponibles, et dans certains cas à des images satellites du Géoportail ou de Google Maps.

Le tableau 3.21 confirme l'exemplarité de toutes les sources d'apprentissage pour tenir compte des spécificités des différentes aires urbaines traitées. Il semble que le type concentré ou dispersé de l'habitat joue plus encore que la taille variable des aires urbaines. La source la plus utilisée est Besançon qui est pourtant la plus petite en population à l'échelle de l'aire urbaine. Cette source permet d'apprendre les classifications sur des aires urbaines plus grandes et plus petites qu'elle, mais aussi sur des aires urbaines aux formes d'habitat concentré mais aussi relativement dispersé. Les trois autres sources trouvent aussi leur intérêt, Tours étant la deuxième la plus utilisée suivie de Lyon puis de Strasbourg. Ces deux plus grandes aires urbaines possèdent l'avantage de fournir plus de labels disponibles pour l'apprentissage, mais elles sont finalement peut-être trop grandes pour apprendre un grand nombre d'aires urbaines.

En termes de modèles, il est intéressant de noter que l'apprentissage de la majorité des aires urbaines est fait en s'adaptant à la distribution des covariables ou à la structure spatiale. Seule La Rochelle et Nantes ont été apprises depuis Besançon sans aucune adaptation, ce qui ne signifie pas nécessairement que ce modèle soit parfaitement adapté. L'adaptation du paramètre spatial β_T n'est pas systématique : il a été ré-appris 18 fois et a été conservé à l'identique du modèle source 28 fois. La famille de modèles la plus utilisée est la famille (M4), suivi de près par la famille (M2). La prépondérance de ces deux modèles semble indiquer que dans la majorité des cas la nécessité du transfert provient de la nécessité de s'adapter à la distribution statistique des classes prises séparément, ou d'adapter simplement le mélange d'un facteur d'échelle indépendant des classes et des variables.

TABLE 3.21 – Aires urbaines composant la base de données des tissus urbains

Région	Aire urbaine (2010)	Population*	Source d'apprentissage	Modèle
Alsace	Strasbourg	761042	-	-
	Mulhouse	281006	Strasbourg	$\beta M1$
Aquitaine	Bordeaux	1127776	Besançon	$M4$
	Pau	240696	Besançon	$\beta M2$
	Bayonne	280538	Besançon	$M4$
Auvergne	Clermont-Ferrand	463891	Besançon	$M4$
Basse-Normandie	Caen	399471	Tours	$M2$
Bourgogne	Dijon	373734	Besançon	$\beta M1$
Bretagne	Rennes	671845	Besançon	$M2$
	Brest	313686	Besançon	$M2$
	Lorient	213189	Besançon	$M4$
	Saint-Brieuc	169638	Besançon	$M4$
Centre	Orléans	419211	Besançon	$M2$
	Tours	477438	-	-
Champagne-Ardenne	Reims	314175	Besançon	$\beta M4$
	Troyes	189811	Besançon	$\beta M1$
Franche-Comté	Besançon	245092	-	-
	Montbéliard	162582	Besançon	$M4$
Haute-Normandie	Rouen	652898	Lyon	$bM1$
	Le Havre	292515	Lyon	$\beta M4$
Languedoc-Roussillon	Perpignan	302 185	Tours	$M3$
	Montpellier	549491	Tours	$M3$
	Nîmes	251 791	Tours	$M2$
Limousin	Limoges	282873	Tours	$\beta M4$
Lorraine	Nancy	434948	Besançon	$\beta M4$
	Metz	389686	Besançon	$\beta M4$
Midi-Pyrénées	Toulouse	1232398	Besançon	$M2$
Nord-Pas-de-Calais	Lille	1158306	Strasbourg	$M2$
	Douai - Lens	542918	Strasbourg	$M3$
	Béthune	367402	Strasbourg	$M2$
	Valenciennes	367941	Strasbourg	$\beta M2$
	Dunkerque	258554	Strasbourg	$\beta M4$

Région	Aire urbaine (2010)	Population*	Source d'apprentissage	Modèle
Pays de la Loire	Nantes	873133	Besançon	$M1$
	Angers	397435	Tours	$\beta M3$
	Le Mans	341080	Besançon	$M4$
	Saint-Nazaire	210574	Besançon	$\beta M4$
Picardie	Amiens	293145	Besançon	$M2$
Poitou-Charentes	Poitiers	252381	Besançon	$M4$
	La Rochelle	204418	Besançon	$M1$
	Angoulême	178496	Besançon	$M4$
Provence-Alpes-Côte d'Azur	Marseille - Aix	1718281	Lyon	$M2$
	Nice	1001295	Lyon	$M2$
	Toulon	606947	Lyon	$M2$
	Avignon	511277	Lyon	$M4$
Rhône-Alpes	Saint-Étienne	508847	Besançon	$M4$
	Chambéry	213251	Besançon	$\beta M4$
	Lyon	2165785	-	-
	Grenoble	669595	Tours	$\beta M3$
	Valence	173973	Besançon	$M2$
	Annecy	217260	Tours	$\beta M2$

* Population 2010 des aires urbaines 2010 (INSEE)

L'influence des sources et des modèles semble liée à un particularisme régional très marqué dans certains cas. Par exemple, toutes les aires urbaines traitées dans le Nord-Pas-de-Calais sont apprises à partir de Strasbourg et principalement à partir des modèles de la famille ($M2$); quasiment toutes celles de la région PACA sont apprises à partir de Lyon et du modèle $M2$; celles du Languedoc-Roussillon sont apprises à partir de Tours et de la famille ($M3$); ou encore toutes celles traitées en Bretagne sont apprises à partir de Besançon. Cette particularité est un fait intéressant à prendre en compte pour renforcer l'apprentissage et sélectionner le modèle le plus pertinent pour chaque nouvelle aire urbaine à traiter.

3.3.3.3 Le rôle des données

Si la majorité des aires urbaines ont pu être apprises correctement grâce à une des sources et un des modèles disponibles, des erreurs sont toutefois observables sur les résultats de classification de plusieurs aires urbaines. Si certaines erreurs sont liées à un problème d'adaptation des modèles, d'autres erreurs récurrentes peuvent être attribuées aux données utilisées qui ne permettent pas dans certains cas de détecter correctement les catégories de tissus urbains.

Détection des centres historiques Parmi les erreurs facilement identifiables, certaines concernent des mailles isolées affectées à la catégorie « centre historique ». Ces erreurs restent relativement rares sur l'ensemble de la base de tissus urbains notamment grâce à la structure de voisinage utilisée lors du décodage du champ de Markov. Elles concernent des mailles isolées qui correspondent en fait à de très grandes emprises industrielles où la surface bâtie est très importante et la part de population non négligeable du fait de l'intégration dans un environnement qui n'est pas exclusivement industriel mais aussi résidentiel. Ces erreurs isolées sont généralement facilement corrigibles par lissage majoritaire dans un voisinage de 3×3 mailles par exemple. Seul un cas plus critique a été identifié sur l'aire urbaine de Montbéliard où une zone de centre historique composée de 3 mailles est identifiée par erreur sur la zone industrielle de Beaulieu à Mandeure.

Détection de l'habitat collectif Les erreurs d'affectation de la population sur les parcelles de taille parfois importante correspondant à des opérations d'aménagement de type grands ensembles causent des problèmes fréquents dans la bonne affectation de ces quartiers à la catégorie « habitat collectif ». En effet, lorsque les quartiers s'étendent sur plusieurs mailles de 200 m de côté, il est fréquent qu'une des mailles possède une valeur très élevée de population tandis que les mailles voisines, même si elles contiennent des immeubles entiers possèdent, elles, des valeurs beaucoup plus faibles voire nulles qui ne sont pas en accord avec la valeur centrale. Cette situation conduit à sous-détecter de façon récurrente les mailles de type grands ensembles déjà rares. Les mailles voisines se retrouvent labellisées en habitat mixte ou en habitat individuel ce qui conduit à un fort morcellement des quartiers d'habitat collectif.

Cette catégorie de tissu urbain apparaît donc comme une des plus difficiles à apprendre et à transférer de par sa rareté et sa forte variabilité en termes de population d'une aire urbaine à l'autre. Lors de la construction de la base de données, trois cas se distinguent : un apprentissage correct détectant les principaux quartiers, une sur-détection intégrant dans la catégorie des mailles faiblement bâties mais proportionnellement fortement peuplées, une sous-détection où seules les mailles extrêmement peuplées sont détectées. Les cas de sur-détection peuvent généralement être facilement corrigés puisque les mailles mal affectées à la catégorie sont généralement des mailles isolées corrigibles par lissage majoritaire dans un voisinage de 3×3 par exemple. Les cas de sous-détection sont plus problématiques puisqu'ils fournissent une simple accroche des plus grands quartiers, qu'il est plus difficile de corriger automatiquement compte tenu des données.

3.3.3.4 Le rôle de la nomenclature de tissus urbains

A côté des difficultés d'apprentissage liées aux données sur lesquelles nous travaillons, il semble important de préciser certaines spécificités observables sur les aires

urbaines, qui sont masquées par la nomenclature volontairement générale et adaptée à la représentation de toutes les aires urbaines.

Par exemple, sur les aires urbaines du littoral méditerranéen et sur celle de Bordeaux, on retrouve plusieurs zones qui sont classées en zones d'activités, mais qui correspondent en fait à des zones à très forte présence de résidences secondaires. Ces zones à localisation quasiment exclusivement littorales sont composées de bâtiments de la taille de logement individuel mais sont associées à une très faible densité de population ce qui conduit à les classer en zones d'activités.

Plusieurs aires du Nord-Pas-de-Calais qui sont intégrées à la base de tissus urbains présentent une forme d'habitat bien particulière qui est l'habitat minier. Corons, cités pavillonnaires ou cités jardins, toutes ces formes d'habitat individuel représentent des tissus urbains spécifiques marqués par la présence de maisons mitoyennes. Ces formes d'habitat apparaissent peu denses en termes de surface bâtie sur les grilles à mailles de 200 mètres de côté puisque ces petites habitations sont construites sur des parcelles étroites et sont associées à des jardins. Étant aussi associées à une faible densité de population ces formes d'habitat se retrouvent généralement dans la catégorie d'habitat individuel groupé. À l'opposé, on retrouve sur des aires urbaines telles que Tours ou Rennes des formes d'habitat individuel de type maisons mitoyennes qui ne sont pas nécessairement associées à une plus haute densité de surface bâtie mais qui sont associées à une plus haute densité de population provoquant le basculement dans la catégorie d'habitat mixte et dense.

Les exemples qui sont mentionnés rappellent que même si nous associons les classifications à des formes générales d'habitat, la nomenclature définie de manière commune sur l'ensemble des aires urbaines est une nomenclature qui a pour vocation de représenter des modulations de densités. Mais il reste des modulations spécifiques liées notamment aux taux d'occupation des logements qui amènent une variabilité des formes qui ne peut être considérée ici.

Conclusion

Ce chapitre nous a permis d'introduire un algorithme original, l'algorithme trN-REM, permettant de classer des données à dépendances spatiales modélisées par champ de Markov discret caché, en utilisant des méthodes de transfert d'apprentissage. Cet algorithme a été employé pour constituer une base de données des tissus urbains sur les 50 plus grandes aires urbaines françaises (exception faite de Paris), tout en limitant le recours à une connaissance spécifique de la composition de chaque aire urbaine puisque seules 4 aires urbaines ont été partiellement labellisées à la main.

L'utilisation de cet algorithme nous a permis de réaliser les diverses classifications

seulement à partir de la connaissance disponible sur quatre aires urbaines : Besançon, Tours, Strasbourg et Lyon. L'adaptation aux distributions des covariables et aux structures spatiales des classes de tissus urbains permet de traiter n'importe quelle aire urbaine française selon une nomenclature commune et sans connaissance particulière sur les nouvelles aires urbaines à classer. Cette méthode de classification a révélé que les liens de distributions des variables de densité de population, de surface bâtie et d'entropie étaient conditionnés par la spécificité des formes de bâties qui semble se retrouver au niveau régional, mais aussi par la taille des aires urbaines.

Au terme de ce chapitre, nous obtenons une représentation qui nous informe sur la composition de ces aires urbaines en termes de tissus urbains qualifiés à partir de niveaux de densité de population, de surface bâtie et d'entropie. Ces regroupements de mailles homogènes constituent un premier apport de structuration spatiale dans les grilles régulières. A partir de maintenant nous quittons l'échelle de la maille pour poursuivre le travail d'exploration des organisations urbaines à l'échelle de la région image homogène du point de la composition morphologique.

Chapitre 4

Représenter les structures urbaines : des tissus urbains aux morphotypes

Sommaire

Introduction	158
4.1 Le modèle chorématique	159
4.1.1 La démarche chorématique	159
4.1.2 Chorèmes urbains	162
4.1.3 Exemples de modèles particuliers	167
4.2 Les morphotypes : Enrichir les tissus urbains grâce au raisonnement spatial	170
4.2.1 Le raisonnement spatial	171
4.2.2 Règles d'identification des morphotypes	178
4.2.3 Évaluation des classifications : interpréter les différences	188
Conclusion	192

Introduction

Dans l'idée que les classifications obtenues dans le chapitre précédent aident à identifier les régions d'image élémentaires qui composent des formes urbaines, il convient maintenant de pouvoir les mettre en relation afin de comprendre et de comparer les logiques qui régissent l'organisation des aires urbaines. Ce chapitre a donc pour objectif d'introduire une représentation des structures urbaines qui sera employée pour observer et interroger la diversité des configurations morphologiques observées parmi les 50 aires urbaines traitées précédemment.

Cette nouvelle représentation se fait dans la continuité des objectifs de modélisation introduits au chapitre 1. Elle correspond à l'identification de nouvelles images classées, cette fois-ci en travaillant à l'échelle meso-géographique de la *région d'image*. Nous souhaitons ainsi proposer une caractérisation des formes urbaines selon une composition en catégories de morphotypes de développement urbain, qui informent sur les logiques d'organisation et sur la morphogenèse urbaine. Ces morphotypes seront identifiés selon le même cahier des charges que celui des tissus urbains, en reprenant des oppositions structurelles de forme, de fonction, de hiérarchie, et surtout de relation.

Ce chapitre est organisé en deux parties. Dans la première partie, nous introduirons la nouvelle nomenclature en morphotypes, construite en ayant recours à la chorématique. Langage de modélisation graphique, la chorématique et les chorèmes servent de légende ou de clés de lecture à la « carte-modèle » (Brunet, 1986) qui fournit une description de la structure et des dynamiques fondamentales de l'espace considéré. Elle sera donc utilisée pour formaliser théoriquement et graphiquement un modèle d'organisation décrivant la structure morphologique des aires urbaines. En permettant d'exprimer les caractéristiques fondamentalement structurantes des espaces urbains, elle constituera la base d'une nouvelle forme de modélisation lisible à l'échelle macro-géographique.

Dans la seconde partie, nous présenterons le volet qualitatif et cartographique de mise en forme des données, réalisé à l'aide des outils du raisonnement spatial permettant d'explicitier le passage des éléments de modélisation théorique fixés par des chorèmes aux modèles particuliers observés sur les grilles régulières. En fixant des règles de représentation spatiale et des règles de relations, nous pourrions faire le lien entre ce langage graphique définissant des représentations « sans échelle, ni orientation » et un langage plus cartographique articulé autour de nos classifications en tissus urbains. Ce lien se fera dans la pratique en définissant un processus de post-classification des images classées en tissus urbains. Précédé d'une phase de simplification de ces images, ce processus aboutira à la requalification des tissus urbains en catégories de morphotypes, par l'application d'un ensemble de règles de décision spatiale faisant intervenir des relations topologiques de partie au tout, d'adjacence ou encore d'inclusion.

4.1 Le modèle chorématique

Dans une démarche de modélisation, le parti pris est qu'il est possible de représenter et de simplifier la complexité de la ville pour fournir une description synthétique des formes et arrangements spatiaux. La réalité que nous souhaitons observer ici est celle des processus structurant l'organisation des espaces urbains. L'objectif est de comprendre comment ces processus organisateurs se mettent en œuvre dans des espaces, tous particuliers. Nous souhaitons affiner la connaissance des espaces urbains dans l'idée de révéler des caractéristiques morphologiques qui puissent servir à la spatialisation de différents enjeux d'aménagements. Dans ce cadre, les modèles théoriques et les différentes caractéristiques d'organisations introduits au chapitre 1 continuent de représenter « une source d'hypothèse de travail à mettre à l'épreuve de la réalité » (Haggett, 1973). Ils représentent le support que nous allons essayer de spatialiser en mettant l'accent sur les relations spatiales, en supposant que ces relations nous permettront d'enrichir la qualification des structures urbaines en affinant les caractéristiques de formes, mais aussi de hiérarchisation et de fonction.

C'est sur cette base que nous formalisons un ensemble de clés de lecture/modèles élémentaires servant à la construction d'un modèle et à l'introduction d'une nomenclature de morphotypes. Nous présenterons dans un premier temps la démarche chorématique, telle qu'elle a été introduite par Brunet (1980), afin de présenter l'intérêt de cette approche pour dégager une représentation générale qui pourra facilement être appliquée à n'importe quel espace urbain. Nous présenterons ensuite les chorèmes urbains permettant de décliner un modèle générique de structure urbaine, qui nous permettra d'aboutir à la constitution d'une nomenclature de morphotypes. Ils seront de plus utilisés pour figurer des modèles plus particuliers permettant de caractériser et de spatialiser toutes sortes d'organisations morphologiques.

4.1.1 La démarche chorématique

Le mot chorème est un néologisme introduit par Brunet (1980) pour désigner un modèle spatial élémentaire qui exprime une forme élémentaire d'organisation de l'espace. En introduisant ce concept, Brunet a introduit une méthode de modélisation et d'analyse spatio-temporelle, la chorématique, propre à la géographie au sens où l'espace est au cœur de cette approche. Basée sur la représentation graphique, cette approche joue un rôle de support de communication situé entre la carte (objet du géographe) et le schéma. La chorématique cherche à représenter et éclairer la compréhension de la variété existant dans les organisations spatiales quelles qu'elles soient. En partant du postulat que chaque organisation spatiale est particulière, Brunet ambitionne de mettre au point une méthode de représentation graphique qui puisse faire figurer cette variété de configurations. L'idée est d'identifier sous une nouvelle forme les régularités

et les différences entre les organisations, vues à travers une représentation des principaux phénomènes marquant ces organisations. L'objectif de la chorématique est ainsi de fournir un nouveau type de modélisation qui permette de passer du modèle théorique, exprimant un phénomène ou mécanisme en jeu, au modèle particulier décrivant plus finement la complexité de chaque espace considéré, résultat de la combinaison de plusieurs modèles élémentaires d'organisations.

Le passage du modèle général au modèle particulier se fait donc à l'aide des chorèmes, ces représentations graphiques, existant en nombre limité, qui traduisent spatialement et graphiquement des éléments simples d'organisation de l'espace. Le chorème n'est ni une carte généralisée ou simplifiée, ni un schéma à proprement parler, il est le figuré graphique (sans échelle, ni orientation) d'un processus qui permet de comprendre et d'interpréter un élément dans les stratégies d'appropriation et d'organisation des territoires. Dans son article de 1986, Brunet expose une table de 28 chorèmes présentés ici en figure 4.1. Cette table, contenant un nombre limité mais non figé de chorèmes, fournit une sorte d'alphabet permettant potentiellement d'écrire toutes les d'organisations spatiales. Un chorème représente ainsi un signe associant une forme (le point, la ligne, l'aire et le réseau) et un signifié c'est à dire un processus mis en jeu pour expliquer l'organisation d'un espace particulier (maillage, quadrillage, attraction, contact, etc.). Ces signifiés peuvent correspondre à des modèles d'organisations spatiales qui peuvent aussi bien relever de stratégies de contrôle territorial que de modélisation de forces physiques (gradient climatique, dissymétrie entraînée par des contraintes topographiques, etc.).

Combiner ces signes permet de construire des mots ou des phrases plus ou moins complexes selon le niveau de spécificité conservé. Mais au-delà de la modélisation par niveaux emboîtés d'un territoire particulier, ce travail de composition en chorèmes effectué sur différents territoires permet de dégager des structures d'organisations récurrentes. Ces structures récurrentes de niveaux intermédiaires qui correspondent à la combinaison de deux ou trois modèles sont qualifiées de chorotypes (Brunet et Théry, 1992). Les chorotypes constituent des modèles plus ou moins complexes qui se révèlent intéressants pour servir de base à la formation de typologies.

Les chorèmes ont donc un double sens, objet graphique et clé d'interprétation à la lecture d'un modèle. Ils permettent de représenter une organisation spatiale sous la forme d'un modèle particulier. Ainsi la construction d'un schéma chorématique implique un va-et-vient entre observation de la réalité et simplification par la pose d'hypothèses pertinentes sur l'organisation et les modèles élémentaires qui en découlent. Comme toute démarche de modélisation, ce va-et-vient est guidé par les objectifs thématiques, par les choix d'abstraction réalisés et donc par le message à faire passer.

Si cette démarche de modélisation graphique a été adoptée par une communauté de géographes pour travailler sur des applications variées (Brunet, 1986), elle a pourtant

	POINT	LIGNE	AIRE	RESEAU
maillage				
	chef-lieu	limite administrative	Etat, région ...	centres, limites et polygones
quadrillage				
	tête de réseau carrefour	voies de communication	aire d desserte irrigation, drainage	réseau
attraction				
	points attirés satellites	lignes d'isotropie orbites	aire d'attraction	liaisons préférentielles
contact				
	point de passage	rupture, interface	aire en contact	base tête de pont
tropisme				
	flux directionnel	ligne de partage	surfaces de tendances	dissymétries
dynamique territoriale				
	évolutions ponctuelles	axes de propagation	aires d' extension	tissu du changement
hiérarchie				
	semis urbain	relation de dépendance limites administratives	sous-ensemble	réseau maillé

FIGURE 4.1 – Table chorématique de Brunet (1986)

été au cœur de nombreux débats et critiques (Lacoste, 1995). La chorématique a été perçue par certains comme une démarche peu scientifique, qui s'éloigne de la « réalité » cartographique et trop dépendante de ce que les théoriciens veulent lui faire dire. Démarche de modélisation graphique et déductive qui ne fait pas appel à une approche quantitative, la chorématique a été critiquée pour l'illusion qu'elle donne « d'une géographie facile qu'on livre avec un "kit" ... sans méthode d'assemblage » (Reymond, 1996). Pour cet auteur ce n'est pas le concept de chorème qui est remis en cause, mais plutôt la méthodologie ou la conceptualisation incomplète de ce nouveau langage cartographique qui est donné sans toutes les règles de grammaire et qui laisse place donc une trop grande part d'interprétation personnelle. Vue comme « une construction provisoire à éprouver, à enrichir » (Brunet, 1987), la chorématique n'en est pourtant qu'à ces débuts. L'intérêt persistant de certains géographes et scientifiques d'autres disciplines (Piveteau et Lardon, 2002; Lardon et Capitaine, 2008; Grataloup et Eckert, 2010) montre que les chorèmes continuent de prouver leur utilité en tant qu'« outil de recherche et de communication » (Brunet, 1986).

Plus que les chorèmes présentés par Brunet, c'est la démarche chorématique qui nous intéresse dans l'objectif d'appliquer cette stratégie de modélisation à la description et à l'interprétation des différences d'organisation des structures urbaines du territoire français. La démarche chorématique nous a paru adaptée dans le sens où elle permet d'explicitier ce que nous considérons comme les modèles élémentaires d'organisation urbaine, communs à la plupart des structures urbaines françaises. Cette approche de composition/décomposition en modèles élémentaires fournit une base méthodologique qui semble pertinente pour expliciter et donner sens aux formes et structures urbaines.

4.1.2 Chorèmes urbains

Les chorèmes que nous définissons ici ont pour rôle de représenter les caractères communs aux structures urbaines, qui une fois qu'ils sont explicités permettent de représenter des situations urbaines variées à travers leur combinaison. Ces chorèmes constituent des éléments de modélisations qui se retrouvent dans d'autres travaux de modélisation chorématique, tels que ceux de Djament-Tran et Grataloup (2010) sur les trajectoires des villes. Le choix des chorèmes retenus repose sur les caractéristiques d'organisation morphologique que nous souhaitons expliquer et mettre en avant, il se fait en tenant compte des données dont nous disposons. Nous nous efforçons de présenter des chorèmes qui permettent d'articuler différents niveaux de modélisation, « de la carte au modèle », et de proposer un enrichissement spatial des classifications en tissus urbains. L'hypothèse est alors que les processus représentés aboutissent à la production d'un maillage, c'est-à-dire une partition de l'espace géographique, que nous cherchons à identifier et qualifier en combinant ces chorèmes.

Nous introduisons dans un premier temps six chorèmes en figure 4.2, permettant

de modéliser les différentes caractéristiques d'organisations morphologiques présentées dans le chapitre 1. Ces figurés élémentaires permettent de représenter de façon générale une forme urbaine métropolitaine marquée par différents processus de développement par densification, extension ou intégration.

On y retrouve la base graphique usuellement employée pour représenter un espace urbain marqué par un modèle de dynamique urbaine en gradient décroissant du centre vers la périphérie. Ce modèle produit une organisation spatiale en aires d'extension emboîtées qui correspondent à un modèle de développement isotrope, qui peut toutefois être déformé par les voies de communication structurant le développement urbain. La prise en compte des voies de communication a un double rôle dans la compréhension des processus d'organisation des espaces urbains. En tant qu'infrastructure physique, elles peuvent guider les choix d'implantation et donc le développement des formes. En tant qu'infrastructure de communication, elles sont le support des interactions plus ou moins fortes entre les lieux composants le systèmes urbains. Elles assurent donc un lien avec la composante spatio-fonctionnelle des organisations, qui n'est pas observée directement dans notre modèle morphologique.

L'intégration de processus différenciation dans les organisations urbaines se fait sous deux formes : une différenciation qui résulte de la sectorisation des fonctions intra-urbaines et une différenciation qui résulte de façon générale des effets de distances et de contacts. La première différenciation correspond aux modèles sectoriels

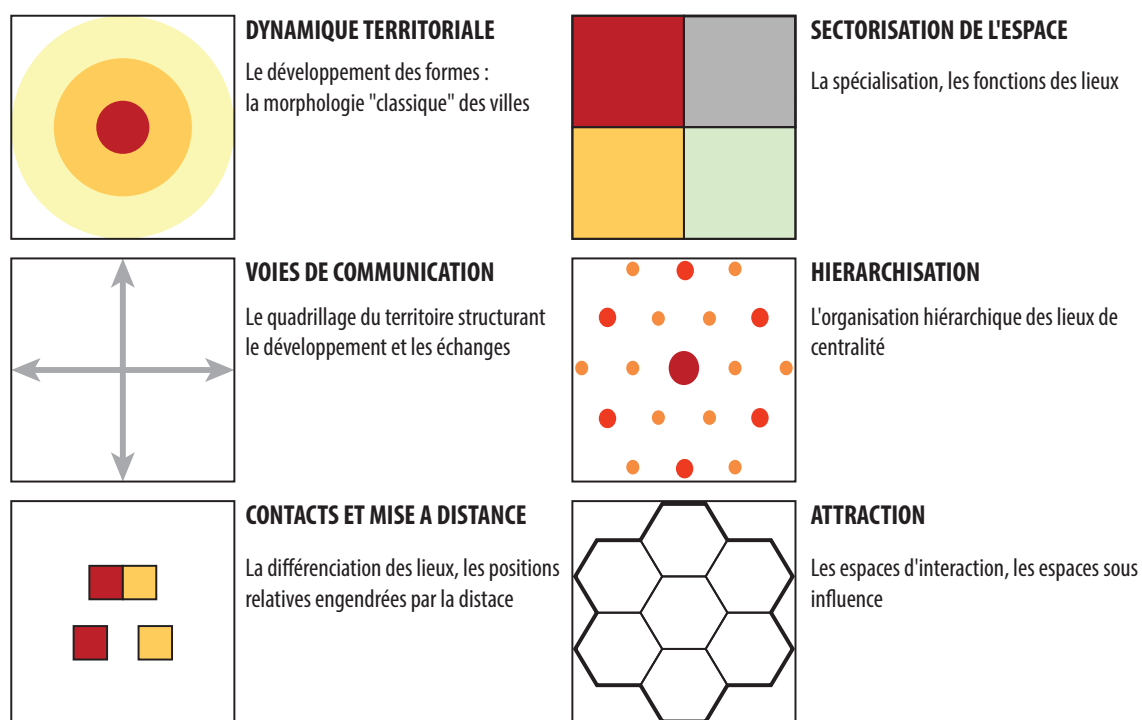


FIGURE 4.2 – Chorèmes urbains

de ville, observée à l'échelle méso-géographique, en tant qu'organisation aux quartiers différenciés selon des fonctions, ou des caractéristiques morphologiques différentes.

En prenant en compte la distance comme clé de lecture graphique, l'idée est de représenter le processus de contact et de mise à distance qui conditionne l'établissement de relations entre des lieux, qui peuvent être observés à une échelle meso ou macro-géographique. La distance engendre des lieux différenciés puisqu'elle constitue une barrière dans l'établissement de relations, un obstacle à franchir, qui d'une part va se traduire par une volonté de se rapprocher. Mais cette barrière est aussi une sécurité qui permet de maintenir la distinction (donc l'identité et le pouvoir) entre des lieux intégrés dans une logique de compétition spatiale. Derrière des notions qualitatives usuellement employées pour décrire des relations de distance entre des paires de lieux comme « loin » ou « proche », « se touchant », « étant séparé par » ou encore « étant relié par », il y a une sémantique qui indique diverses positions, donc diverses formes de relations et d'organisations. Ce chorème urbain trouve son inspiration dans le domaine du raisonnement spatial, ou dans des travaux tels que ceux de [Lardon et Capitaine \(2008\)](#) sur la construction de modèles d'exploitations agricoles.

Les processus de hiérarchisation sont intégrés graphiquement par deux chorèmes qui représentent une version de la théorie des lieux centraux de [Christaller \(1933\)](#). La théorie des lieux centraux qui est utilisée ici fournit une modélisation simple, utilisant peu d'hypothèses pour représenter deux informations différentes mais tout aussi intéressantes pour notre modélisation. Elle permet de représenter d'une part un semi de lieux centraux qui matérialisent des objets aux formes et aux caractéristiques particulières en termes de fonction, de population et de type de bâti. D'autre part elle permet de représenter des zones d'attraction emboîtées qui, sans matérialiser d'objet particulier, révèle les objets (centraux ou non) soumis à l'influence d'un centre. Ces aires représentent des espaces privilégiés d'interactions et d'échanges, qui permettent de définir un dernier niveau de hiérarchie composé de tous les objets contenus dans le périmètre, mais sans caractéristiques visibles de centralité. Nous reprenons donc ce modèle graphique pour mettre en avant de façon générique les processus de hiérarchisation des lieux. Mais nous nuancerons dans la suite le remplissage régulier de l'espace associé à ce modèle graphique qui est rendu inopérant par différentes contraintes topographiques et historiques qui traduisent l'anisotropie d'organisation de la majorité des objets urbains.

Ces six clés de lecture correspondent ainsi aux processus que nous avons choisi de modéliser pour interpréter les différences marquant entre les structures urbaines. Elles correspondent aux processus qui semblent indispensables pour comprendre les logiques de développement urbain et de mises en relations des espaces intra-urbains ([Haggett, 1973](#)). Bien sûr elles pourraient être raffinées et complétées par d'autres processus structurant les organisations urbaines. Mais leur choix repose sur la volonté de faire évoluer les classifications des tissus urbains qui nous servent de base d'observation

des structures urbaines. Elles représentent donc des processus théoriques qui peuvent en pratique être quantifiés à partir des données dont nous disposons, qu'elles soient relatives aux classifications.

La combinaison de ces chorèmes nous permet donc de définir un modèle graphique abstrait de structure urbaine. La simplification maximale des formes spatialisées selon des hypothèses de localisation régulière et géométrique aboutit à la formalisation d'un modèle générique donné en figure 4.3. Nous introduisons cette figure pour présenter les différents types d'objets urbains composant les formes urbaines métropolitaines, que nous qualifions de morphotypes. Ces morphotypes correspondent aux catégories d'espaces bâtis qui peuvent être identifiés comme étant le résultat des différents processus d'organisation représentés par les chorèmes de la figure 4.2. Sur cette figure la spatialisation de ces morphotypes correspond à une vision générale et simplifiée des organisations urbaines où l'on s'abstrait des contraintes spatiales, qu'elles soient topographiques ou historiques. Nous verrons dans la suite que le caractère générique de cette organisation spatiale peut être déformé en tenant compte de ces contraintes, mais sans modifier la nomenclature des morphotypes présentée (cf. figures ou 4.5 4.6).

Dans ce modèle, les processus d'organisation matérialisent 12 catégories de morphotypes que nous chercherons à identifier dans la suite à partir des classifications en tissus urbains. Trois niveaux de centres sont déterminés par leur taille, leur densité de bâtiment et de population. Ces centres sont les pôles dominants, plus ou moins

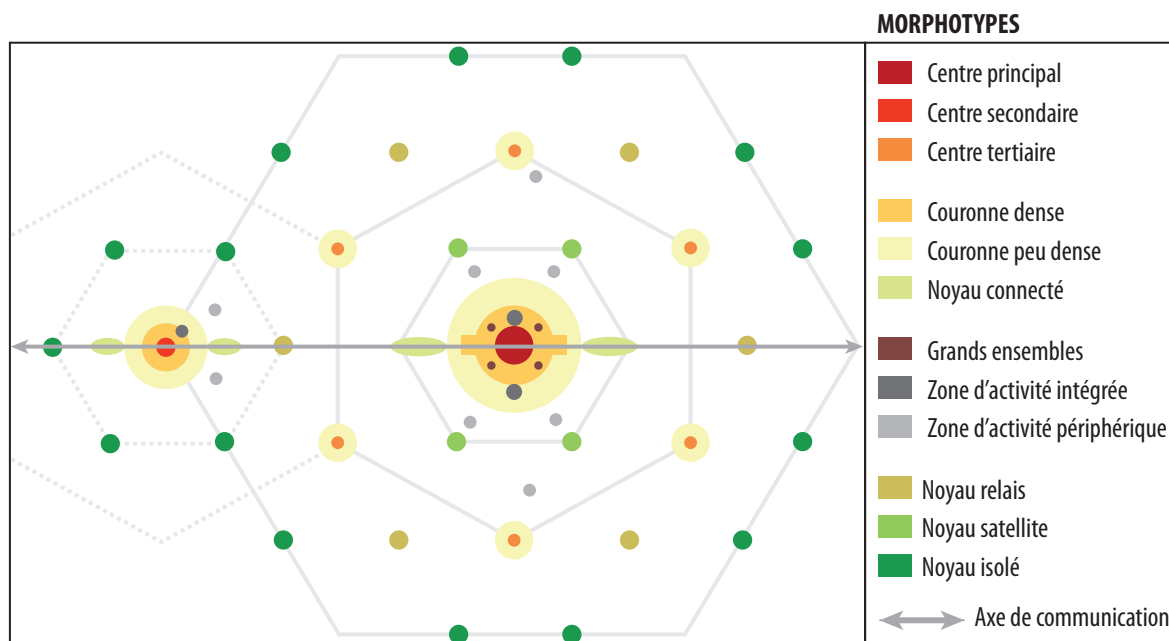


FIGURE 4.3 – Représentation chorématique générique d'une forme urbaine métropolitaine (sans prise en compte de contraintes spatiales influençant les processus de hiérarchisation)

hiérarchisés entre eux, des systèmes définis par les aires urbaines. Ils représentent le cœur d'agglomérations composées selon des logiques d'organisation radioconcentrique et sectorielle.

Ces logiques génèrent des agglomérations où plusieurs types de formes déterminées par leurs caractéristiques morphologiques et leurs fonctions se positionnent de façon adjacente. Ces agglomérations sont marquées par un développement en couronnes, qui sont distinguées en deux catégories : les couronnes denses qui sont les premières extensions spatiales des centres de plus haut niveau de densité ; et les couronnes peu denses qui sont les deuxième extensions des centres de plus haut niveau de densité, mais aussi les premières extensions des centres tertiaires de plus petite taille. Elles sont aussi marquées par une logique de sectorisation qui laisse apparaître des espaces à la morphologie spécifique comme les grands ensembles, et des espaces aux fonctions spécifiques comme les zones d'activités. Ces zones d'activités sont elles-mêmes distinguées par leur position intégrée ou périphérique à une agglomération, qui leur confère un rôle distinct dans le développement urbain, pour des questions de consommation d'espace et de périurbanisation des activités.

Comme représenté sur l'agglomération principale du modèle, la présence d'infrastructures de transport déforme les couronnes par des logiques de localisation préférentielle à leur proximité. Les infrastructures de transport guident le développement urbain orienté par des processus d'étalement depuis le centre. Mais elles guident aussi des processus d'intégration d'espaces périphériques anciennement indépendants morphologiquement. Ces espaces sont matérialisés par la catégorie des noyaux connectés. Ces noyaux représentent une instance de développement fréquente dans les formes métropolitaines marquées par la coalescence. Cette coalescence, dans le cas le plus extrême, peut prendre la forme d'une conurbation où toutes les agglomérations majeures du système sont connectées.

La hiérarchisation des lieux peut aussi être observée par l'analyse des aires d'influence qui permettent de qualifier le bas de la hiérarchie urbaine. Les derniers niveaux de la hiérarchie correspondent aux formes déconnectées des lieux centraux. Dans ces aires d'influence, on définit trois catégories de forme : les noyaux relais, les noyaux satellites et les noyaux isolés. Ces différents types de noyaux sont distingués par leur position et leur fonction. Les noyaux relais correspondent à des zones résidentielles associées à des zones d'activité. La présence de ces zones d'activité leur confère un pouvoir structurant au sein des périphéries. Les noyaux satellites sont des espaces associés majoritairement à des fonctions résidentielles, mais qui présentent la particularité de se situer à proximité d'agglomérations centrales, bien qu'ils en soient déconnectés. Ils sont distingués des noyaux isolés, qui eux aussi sont associés à des fonctions majoritairement résidentielles, mais se situent à des positions plus éloignées des agglomérations. On peut supposer que cette différence de position génère des caractéristiques différentes en termes de développement et de croissance urbaine.

Ce modèle définit donc la structure d'une forme urbaine métropolitaine. Cette forme très hétérogène est composée d'objets fragmentés spatialement, qui possèdent des fonctions différenciées d'habitat et d'activité. Elle est structurée par des mouvements et des échanges qui assurent leur cohérence en tant que système intégré et hiérarchisé. Elle existe donc ici dans les limites des espaces définis par les aires urbaines, qui nous assurent cette cohérence fonctionnelle impossible à attraper par des caractéristiques morphologiques. Ici, l'accent est mis sur la présentation d'une nomenclature de formes de développement que nous chercherons à spatialiser de manière plus particulière pour chacune des aires urbaines analysées sur les grilles régulières. Pour les spatialiser à partir de données observées, nous aurons besoin de prendre en compte des caractéristiques plus spécifiques à chacune de ces aires urbaines, qui ne se présentent jamais comme ce modèle théorique régulièrement organisé.

4.1.3 Exemples de modèles particuliers

Dans l'objectif de pouvoir modéliser et comparer différentes structures urbaines il semble difficile d'ignorer certains éléments relatifs aux caractéristiques de site des agglomérations qui jouent un rôle important dans la compréhension des logiques de développement urbain. L'espace n'est pas isotrope et des contraintes topographiques, aussi bien qu'anthropiques, peuvent déformer la représentation graphique de l'organisation parfaitement géométrique de la forme métropolitaine présentée précédemment. Pour prendre en compte la réalité des organisations urbaines qui ne correspond pas à des anneaux concentriques parfaitement emboîtés, il est nécessaire d'introduire des caractéristiques permettant de représenter de façon générique des contraintes spatiales d'organisation. La figure 4.4 présente donc trois nouveaux chorèmes qui permettent de spatialiser les particularités d'organisation observables sur les différentes aires urbaines.

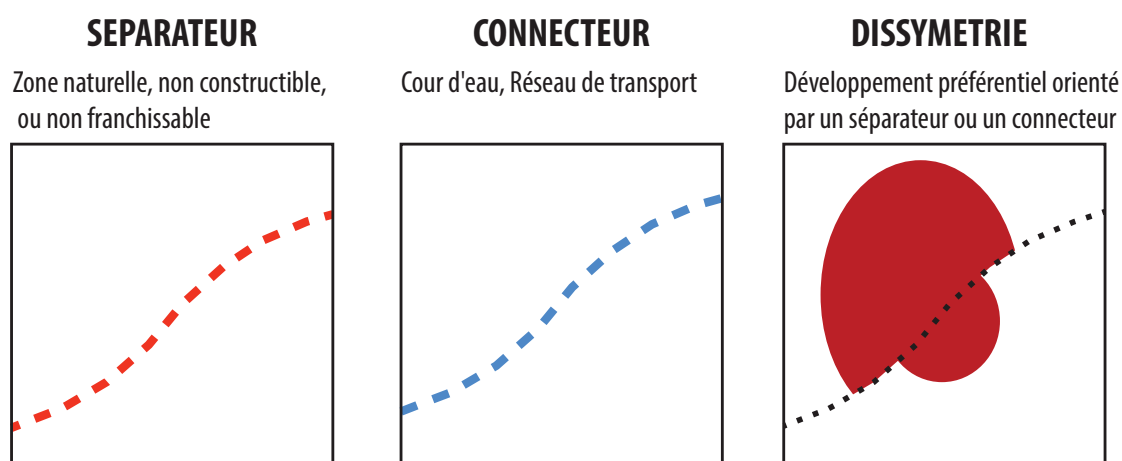


FIGURE 4.4 – Contraintes spatiales générant des particularités d'organisation

Ces chorèmes ont pour vocation de représenter les éléments qui matérialisent des axes de dissymétrie et de rupture entraînant des développements inégaux de certaines formes urbaines. Ce qui est qualifié d'axe ici dénote des objets et phénomènes de nature différente que nous distinguons en deux catégories. La première catégorie regroupe les objets qui jouent un rôle de séparateur, de limitation dans le développement urbain. L'axe de rupture peut être alors être la limite du front d'urbanisation bloqué par la présence d'une zone non constructible ou non franchissable à cause de contraintes liées par exemple à l'environnement physique et à la topographie, ou à des interdictions de construction. La deuxième catégorie regroupe à l'opposé les objets jouant un rôle de connecteur, bien qu'il marque une rupture dans le développement. Classiquement l'axe de dissymétrie connecteur peut être le cours d'eau traversant une agglomération qui entraîne un développement inégal et préférentiel sur l'une des rives. Mais il peut aussi être une frontière territoriale, qui peut assurer un rôle de connecteur et de fixateur du développement (typiquement les agglomérations frontalières de l'Est français). Les frontières territoriales peuvent toutefois être interprétées de manière opposée comme des séparateurs qui marquent des limites nettes dans les fronts d'urbanisation. Ces deux chorèmes sont donc liés au troisième qui matérialise les développements de formes dissymétriques entraînés par la présence de ces axes.

Ces éléments de particularité sont représentés volontairement de manière assez générale, puisque l'objectif est toujours de construire des modèles. L'idée est d'identifier des axes de dissymétrie ou des contraintes de développement qui se retrouvent de manière similaire dans certaines agglomérations rendues invariantes par rotation. En utilisant ces chorèmes, on peut donc proposer des modèles plus particuliers d'organisation, comme ceux des aires urbaines de Besançon (figure 4.5) et de Lyon (figure 4.6). Ces deux figures donnent une vision des éléments les plus marquants de la structure de ces aires urbaines.

L'aire urbaine de Besançon s'organise principalement autour de l'unique agglomération de Besançon. L'agglomération n'est plus représentée de manière circulaire, mais de manière elliptique. Le grand axe de cette ellipse coïncide avec un des axes de transport structurant, lui-même aligné sur l'axe de dissymétrie généré par la double présence du Doubs et de contraintes topographiques. Les axes de transport jouent un rôle structurant sur le développement de l'agglomération principale et au-delà. Ils permettent de comprendre la localisation préférentielle des zones d'activité et noyaux connectés à l'agglomération, mais aussi celle des noyaux relais, associés à des zones d'activité hors agglomération. La présence des différentes contraintes topographiques explique aussi le développement de l'aire urbaine majoritairement orienté sur les plateaux du quart nord-ouest.

L'aire urbaine de Lyon, dont le développement est beaucoup plus complexe que celui de Besançon, s'organise elle autour d'une agglomération principale (celle de Lyon) qui est relayée par plusieurs agglomérations de plus petites tailles. On retrouve ici un

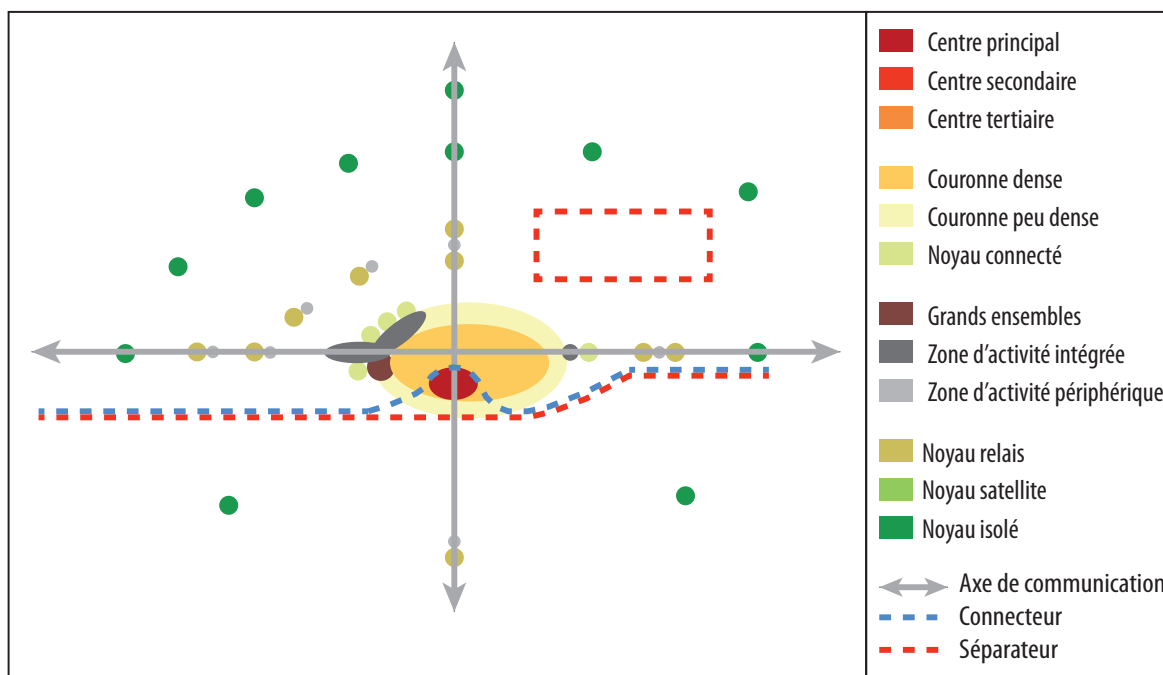


FIGURE 4.5 – Modèle chorématique de l'aire urbaine de Besançon

espace beaucoup plus hiérarchisé avec au Nord l'agglomération de Villefranche-sur-Saône en deuxième position, au Sud la plus petite agglomération centrée sur Givors, toutes les deux morphologiquement connectées à celle de Lyon. Plus encore qu'à Besançon, les axes de transport sont le support d'un développement en continu entre les agglomérations. Le Rhône matérialise un axe de dissymétrie qui se traduit par un développement beaucoup plus industriel et marqué par la présence d'agglomérations indépendantes à l'Est. On retrouve par exemple les agglomérations d'Ambérieu-en-Bugey ou de Bourgoin-Jallieu, et surtout la ville nouvelle de l'Isle-d'Abeau. L'Ouest est lui beaucoup plus résidentiel et avec une densité plus faible, mais il est aussi marqué par la coalescence des formes.

Ces deux modèles constituent donc des exemples de ce qui peut être obtenu avec ce mode de représentation chorématique. Cette façon de procéder fournit une représentation graphique qui facilite grandement la comparaison des structures urbaines étudiées par l'observation des modèles élémentaires mis en jeu dans la production de ces espaces. Au-delà de la représentation, la chorématique fournit une démarche d'interprétation qui peut être utilisée pour « faire parler les formes » et formaliser des nomenclatures que ce soit à l'échelle meso-géographique dans le cadre de ce chapitre, ou à l'échelle macro-géographique dans la suite des travaux.

La question qui se pose maintenant est celle de l'opérationnalisation de règles d'analyse chorématique qui nous permettraient d'obtenir des modèles d'organisation particuliers pour les 50 aires urbaines de notre échantillon d'étude. Dans ce qui suit, nous

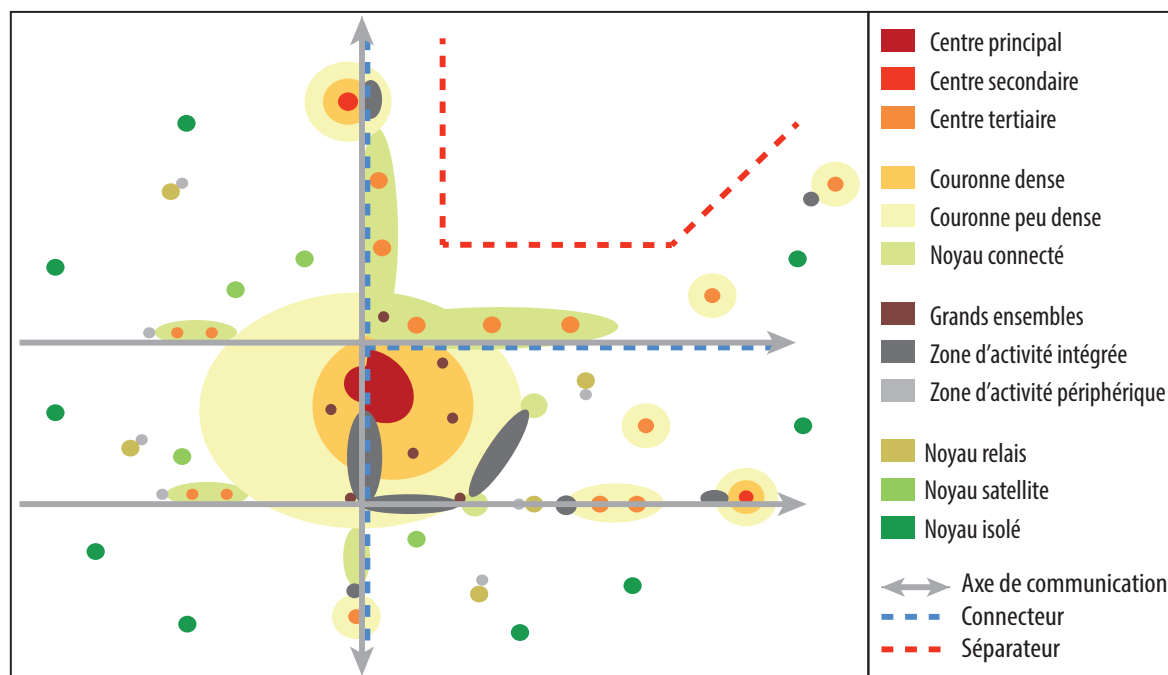


FIGURE 4.6 – Modèle chorématique de l'aire urbaine de Lyon

proposons d'effectuer ce travail en construisant de façon automatique des images classées selon la nomenclature en morphotypes. Nous verrons donc dans la suite les outils d'analyse nécessaires pour faire correspondre ces représentations graphiques simplifiées à des représentations cartographiques produites de façon systématique en utilisant ces règles d'analyse chorématique.

4.2 Les morphotypes : Enrichir les tissus urbains grâce au raisonnement spatial

Les chorèmes urbains présentés dans la section précédente fournissent un moyen d'interpréter les processus de spatialisation en jeu dans la formation d'une structure urbaine. La démarche chorématique nous a permis de proposer des modèles de l'organisation morphologique des aires urbaines, elle nous a permis de construire une nomenclature de morphotypes composant ces modèles. L'objectif est donc maintenant de pouvoir revenir aux données pour identifier de façon systématique ces modèles à partir des images classées de chaque aire urbaine.

La démarche de modélisation se poursuit donc ici en reprenant les classifications en tissus urbains et en se plaçant à l'échelle d'étude de la région d'image pour observer la diversité des structures urbaines. En se référant à la modélisation chorématique et à la nomenclature de morphotypes, nous proposons d'enrichir la caractérisation de ces tissus urbains en intégrant des informations positionnelles et relationnelles qui permettent de

matérialiser l'organisation des espaces urbains.

Nous proposons ici de mettre au point une méthode d'assemblage, parfois considérée comme manquante à la chorématique (Reynaud, 1981), pour obtenir de façon automatique des images classées contenant les éléments de notre modélisation chorématique. L'idée est alors d'utiliser des modèles de logique formelle issus du raisonnement spatial pour déterminer de façon automatique les processus de combinaison des chorèmes urbains. Le raisonnement spatial représente ainsi la boîte à outils permettant d'inférer et d'instancier des modèles d'organisation observés à partir des données sur grille régulière. Nous présenterons dans un premier temps un état de l'art des modèles de raisonnement spatial, qui nous servira à introduire notre modèle de représentation et de raisonnement sur les classifications en tissus urbains. En se basant sur ce modèle de raisonnement nous proposons des règles de décision, appliquées sur une version simplifiée des classifications en tissus urbains. Elles permettront d'obtenir des classifications selon la nomenclature des morphotypes présentée précédemment, qui seront ensuite évaluées à partir de deux nouveaux exemples d'aires urbaines tirés de l'échantillon d'étude.

4.2.1 Le raisonnement spatial

Le raisonnement spatial est le problème de la représentation et de la formalisation de la connaissance sur la position d'entités spatiales et sur les relations qu'elles entretiennent. C'est un problème qui a été formalisé relativement récemment dans le domaine de l'intelligence artificielle (Stock, 1997). Le raisonnement spatial est en effet un problème de modélisation mathématique présent dans diverses tâches d'informatique et d'algorithmique. Il est appliqué dans des tâches de navigation en intelligence artificielle et robotique (Kuipers, 1978). Il constitue une composante intégrée à diverses tâches liées aux systèmes d'information géographique comme l'interrogation de données spatiales par requêtes (Egenhofer et Franzosa, 1991; Clementini et al., 1994), ou la généralisation de données cartographiques (Steiniger et Weibel, 2007). Il est utilisé dans divers application de traitement d'images relatives à la reconnaissance de formes (Lee et Hsu, 1992), à l'imagerie médicale (Bloch, 1999), ou à la télédétection (Inglada et Michel, 2009).

Concrètement la représentation des relations spatiales consiste à définir un ensemble de relations considérées comme fondamentales pour décrire les arrangements spatiaux possibles entre entités. La complexité de la tâche tient à la sélection et à la transposition en concept mathématique ou du moins informatisable d'un riche lexique décrivant les relations spatiales en langue naturelle. La variété des relations possibles a conduit à la formalisation de plusieurs ensembles de relations fondamentales (cf. les ouvrages d'état de l'art de Stock (1997); Cohn et Renz (2008); Steiniger et Weibel (2007)), parmi lesquels trois grandes catégories de relations apparaissent de façon récurrente : les relations topologiques, les relations de distance et les relations d'orientation.

Plusieurs approches de modélisation sont possibles pour représenter des relations spatiales : le raisonnement qualitatif basé sur la construction d'ontologies à partir de représentations symboliques et de logiques formelles (Cohn et Renz, 2008) ; et le raisonnement spatial quantitatif lié aux relations métriques et à l'emploi d'approche probabiliste ou de logique floue pour tenir compte de l'imprécision de certaines mesures ou représentations des entités (Colliot, 2003). Ces deux branches se distinguent généralement par les domaines d'application dans lesquels elles ont été développées. Les approches qualitatives sont généralement liées aux SIG et aux données vectorielles tandis que les approches quantitatives se retrouvent plus souvent dans des tâches de traitement d'images.

Dans ce chapitre nous nous intéressons particulièrement au raisonnement spatial qualitatif et aux relations spatiales topologiques pour enrichir la caractérisation des zones géographiques précédemment identifiées. Mais dans le chapitre suivant nous emploierons dans une moindre mesure un raisonnement spatial plus quantitatif nécessaire pour parfaire la modélisation des structures urbaines. Dans les deux cas, les ressources méthodologiques utilisées en pratique sont des transformations issues de la morphologie mathématique, cadre puissant et flexible de raisonnement spatial (Bloch, 2002). La suite de cette section est donc consacrée à la présentation des opérateurs de base de morphologie mathématique, puis à la présentation du formalisme de raisonnement spatial retenu, adapté au traitement des données à l'aide de ces opérateurs.

4.2.1.1 Opérateurs de morphologie mathématique

La morphologie mathématique est une technique mathématique étudiant les formes, leurs structurations et leurs transformations en traitement d'images. Cette théorie est basée sur la théorie des ensembles et des treillis, la topologie et l'algèbre (Serra, 1982). Par l'emploi d'un élément structurant il est possible de dégager les structures et formes les plus significatives d'une image en analysant les liens entre un objet et son environnement immédiat. La morphologie mathématique possède tous les éléments pour se positionner comme un outil puissant de raisonnement spatial. Bloch (2002) montre en effet que la morphologie mathématique constitue un cadre unifié pour définir des relations spatiales en environnement quantitatif, qualitatif ou même en environnement semi-quantitatif reposant sur une logique floue. Nous introduisons ici les opérateurs basiques de morphologie mathématique dont les définitions sont tirées de l'ouvrage de Serra (1982) et sont données dans le cas binaire.

Soit \mathcal{S} un espace dans \mathbb{Z}^2 support des images binaires considérées. Soit B le sous-ensemble de \mathcal{S} appelé élément structurant. L'élément structurant joue un rôle clé dans la définition des opérateurs morphologiques et dans leurs applications qui dépendront de sa forme. Il existe en effet plusieurs façons de définir un élément structurant : par sa géométrie (carré, cercle, rectangle), par une liste de points (forme arbitraire) ou par une distance (notamment intéressant pour définir des transformées géodésiques).

En pratique nous utiliserons un élément structurant carré ou en croix (cf. figure 4.7), similaire aux systèmes de voisinages de premier ordre ou de second ordre contenant les 4 ou 8 plus proches voisins utilisés dans le chapitre 3.



Croix (voisinage d'ordre 1) Carré (voisinage d'ordre 2)

FIGURE 4.7 – Éléments structurants utilisés

L'érosion $E_B(X)$ et la dilatation $D_B(X)$ d'un ensemble X par un élément structurant B sont définies par :

$$E_B(X) = \{x \in S \mid B_x \cap X\} \quad (4.1)$$

$$D_B(X) = \{x \in S \mid B_x \cap X \neq \emptyset\} \quad (4.2)$$

où B_x dénote l'élément structurant B centré en x . L'effet de ces transformations sur l'ensemble X est, pour l'érosion, d'éliminer les parties connexes plus petites que B ou de déconnecter des parties reliées par des « caps » plus grands que B par exemple. À l'inverse, la dilatation va boucher les trous plus petits que B ou ressouder des parties proches par exemple. Ces deux opérateurs fondamentaux permettent de définir, par composition, d'autres transformations comme l'ouverture $O_B(X)$ et la fermeture $F_B(X)$:

$$O_B(X) = D_B[E_B(X)] \quad (4.3)$$

$$F_B(X) = E_B[D_B(X)] \quad (4.4)$$

L'application de ces transformations aura un effet de lissage sur la forme de l'ensemble X qui opère de façon à peu près similaire aux érosions et dilations mais en conservant une taille plus proche de la taille initiale de X .

Nombreuses sont les transformations qui peuvent être définies à partir de ces opérateurs de bases, mais toute la richesse des transformées de morphologie mathématique ne sera pas nécessaire ici. Ces simples transformations associées à des opérateurs logiques introduits dans la suite seront suffisantes pour décrire les relations spatiales potentiellement observables sur les classifications en tissus urbains.

4.2.1.2 Représentation de l'espace géographique

Avant d'aborder la question de la définition des relations spatiales, il est nécessaire d'aborder la question de la définition des objets composant l'espace géographique. Les modèles traditionnellement employés en raisonnement spatial qualitatif définissent les entités spatiales à partir d'une représentation ponctuelle ou d'une représentation

surfacique. S'il est plus facile de travailler sur des modélisations ponctuelles comme c'est le cas dans les travaux de Egenhofer et Franzosa (1991), Cohn et Renz (2008) notent eux qu'il est plus naturel de travailler sur des entités surfaciques pour représenter des objets physiques toujours perçus sous forme surfacique. Ces entités surfaciques étant toujours associables à une représentation ponctuelle si besoin, la majorité des modèles utilisent une représentation surfacique à partir d'objets élémentaires notés régions.

Assimilables en général à des ensembles en topologie ou dans la théorie des ensembles, ces régions ne sont toutefois pas définies de la même façon dans les diverses ontologies formalisant des raisonnements et relations spatiales. Les points de différences les plus notables concernent le choix de considérer ou non des régions nulles (ou vides), le choix de considérer des régions de dimensions différentes, et le choix de considérer ou non les régions comme des ensembles nécessairement connexes (Cohn et Renz, 2008).

Nous détaillons ici la définition d'une région utilisée dans le modèle RCC (*Region Connection Calculus*) de Cohn et al. (1997). Ce modèle repose sur le choix délibéré de ne pas définir les régions selon toutes les propriétés topologiques (ouvert, fermé, intérieur, frontière, etc.) qui ajoutent des distinctions non nécessaires pour formaliser un raisonnement spatial qualitatif pertinent. Les régions sur lesquelles la théorie est formée sont alors définies selon les propriétés suivantes :

1. Une région est une entité spatiale de dimension arbitraire (1-D, 2-D, 3-D, etc.), avec la seule contrainte que toutes les régions étudiées aient la même dimension.
2. Une région ne peut être nulle.
3. Une région peut être composée d'entités spatialement distinctes.
4. Une région peut posséder des trous.
5. Deux régions sont identiques si elles possèdent la même adhérence (cf. définition en annexe, volume II)

La formalisation de la théorie RCC repose donc sur la définition de cette entité Région, mais aussi sur les définitions des entités Espace Physique (ou région univers *u*) représentant l'espace lui-même et sur l'entité Nul représentant les entités spatiales sans extension physique. Dans la suite, nous proposerons une représentation de l'espace des aires urbaines reprenant en partie cette approche.

4.2.1.3 Relations topologiques

Les relations topologiques modélisent des expressions telles que « à côté de », « à l'intérieur de » ou « se croise ». Ces relations fondamentales en raisonnement spatial sont sans doute les plus étudiées (Balbiani et Muller, 2000) et constituent le cœur de nombreuses approches qualitatives. La plupart de ces approches se basent sur la théorie de la méréologie (Simons, 1987) traitant des relations entre la partie et le tout, autrement dit les relations d'inclusion, et sur son extension la méréotopologie (Varzi, 1996) prenant en compte les notions de connexités.

L'ensemble des relations possibles entre régions est généralement défini par l'énumération de relations restreintes aux relations binaires qui s'expriment à partir de tous les résultats possibles d'union et d'intersection entre paires de régions. Parmi les travaux faisant appel à ce type de représentation se retrouvent les travaux d'Allen (1983) qui modélisent les JEPD (*Jointly Exhaustive and Pairwise Disjoint*) relations topologiques sur des intervalles temporels mais qui ont largement été repris en raisonnement spatial; les travaux de Clarke (1981) définissant un ensemble de 44 relations topologiques selon une approche de méréotopologie; ou les travaux de Cohn et al. (1997) se basant aussi sur une approche méréotopologie pour définir sur le modèle RCC (qui sera présenté ici). Les travaux de Egenhofer et Franzosa (1991), représentatifs des formalismes mêlant représentations ponctuelles et surfaciques, ont notamment conduit à la définition du modèle des 9-intersections utilisé pour l'implémentation de requêtes dans divers SIG. Les travaux de Clementini et Felice (1997) ou Cristani et al. (2000) basés sur le formalisme « jaune d'œuf » (*egg-yolk*) proposent un raisonnement sur des régions aux frontières indéfinies qui peuvent donc être représentées de manière linéaire ou surfacique. Il serait possible de mentionner encore bien d'autres formalismes, mais nous renvoyons à des ouvrages tels que ceux de Stock (1997) et Aiello et al. (2007) pour un état de l'art plus complet.

Nous continuons ici la présentation de la théorie RCC qui est suffisamment générale et flexible pour être mise en œuvre dans notre cadre de travail. La base de cette théorie est la réinterprétation de la relation de connexion (relation réflexive et symétrique) $C(x,y)$ définie par Clarke (1981). Cette relation qui représente « x est connecté à y » est interprétée en termes de modélisation par « les fermetures topologiques de x et y partagent au moins un point ». Elle sert de primitive à la formalisation de la théorie RCC, et permet de définir dix relations supplémentaires appliquées à des entités Région, dont la définition¹ et l'interprétation sont données dans le tableau 4.1. Cohn et al. (1997) prouvent que parmi ces relations, seules huit sont suffisantes pour représenter des relations topologiques. Ces relations sont illustrées en figure 4.8, certaines apparaissant deux fois à cause de leur asymétrie. Ces relations constituent la version basique de la théorie RCC, notée dans ce cas RCC-8 ou RC-{C}.

Mais les auteurs présentent plusieurs extensions de leur raisonnement spatial, no-

1. La définition des connecteurs logiques est donnée dans la table des notation en début de manuscrit.

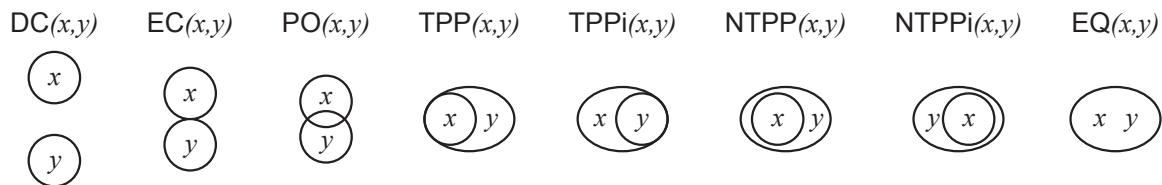


FIGURE 4.8 – Illustration des relations topologiques du modèle RCC-8

TABLE 4.1 – Relations topologiques définissables à partir de la primitive \mathbf{C}

Relation	Définition logique	Interprétation
$\mathbf{DC}(x,y)$	$\neg \mathbf{C}(x,y)$	x et y sont déconnectés (<i>disconnected</i>)
$\mathbf{P}(x,y)$	$\forall z[\mathbf{C}(z,x) \rightarrow \mathbf{C}(z,y)]$	x est une partie de y
$\mathbf{PP}(x,y)$	$\mathbf{P}(x,y) \wedge \neg \mathbf{P}(y,x)$	x est une partie propre de y
$\mathbf{EQ}(x,y)$	$\mathbf{P}(x,y) \wedge \mathbf{P}(y,x)$	x et y sont égales
$\mathbf{O}(x,y)$	$\exists z[\mathbf{P}(z,x) \wedge \mathbf{P}(z,y)]$	x et y se chevauchent (<i>overlap</i>)
$\mathbf{DR}(x,y)$	$\neg \mathbf{O}(x,y)$	x et y sont distinctes (<i>discrete from</i>)
$\mathbf{PO}(x,y)$	$\mathbf{O}(x,y) \wedge \neg \mathbf{P}(x,y) \wedge \neg \mathbf{P}(y,x)$	x et y se chevauchent partiellement
$\mathbf{EC}(x,y)$	$\mathbf{C}(x,y) \wedge \neg \mathbf{O}(x,y)$	x et y sont connectées extérieurement
$\mathbf{TPP}(x,y)$	$\mathbf{PP}(x,y) \wedge \exists z[\mathbf{EC}(z,x) \wedge \mathbf{EC}(z,y)]$	x est une partie propre tangentielle de y
$\mathbf{NTPP}(x,y)$	$\mathbf{PP}(x,y) \wedge \neg \exists z[\mathbf{EC}(z,x) \wedge \mathbf{EC}(z,y)]$	x est une partie propre non tangentielle de y

tamment la version $\mathbf{RC}\{-\mathbf{C}, \mathcal{B}\}$ avec l'ajout d'une liste \mathcal{B} de fonctions booléennes en primitive. Pour enrichir la caractérisation des régions et de leurs relations les fonctions booléennes suivantes sont considérées : **sum**(x,y), la somme des régions x et y (noté +); **prod**(x,y), le produit de x et y (noté *); **diff**(x,y), la différence de x et y (noté -); **compl**(x), le complémentaire de x . Parmi les caractéristiques de forme qu'il est alors possible de définir, retenons qu'il est possible d'introduire un prédicat testant la connexité d'une région :

$$\mathbf{CON}(x) \stackrel{\text{def}}{=} \forall yz[\mathbf{sum}(y,z) = x \rightarrow \mathbf{C}(y,z)] \quad (4.5)$$

Ce prédicat nous servira dans la pratique à faire la distinction entre une région connexe et une région non connexe qui peut être un ensemble de parties connexes. Ces fonctions booléennes, tout comme ce prédicat, trouvent des équivalents directs en traitement d'images. En effet, elles correspondent aux simples opérations d'arithmétiques sur images binaires. De même, la propriété de connexité est aussi définie en traitement d'image (pour la recherche de composante connexe) selon un système de voisinage de premier ou de second ordre équivalent à l'élément structurant B en morphologie mathématique.

Une extension supplémentaire de la théorie RCC repose sur l'ajout de la primitive **conv**(x) définissant la version $\mathbf{RC}\{-\mathbf{C}, \mathcal{B}, \mathbf{conv}\}$. Cette primitive qui dénote l'enveloppe convexe de la région x est définie de façon axiomatique dans (Cohn et al., 1997), et trouve aussi un équivalent direct en traitement d'image. Elle permet de définir de nouvelles relations représentant les notions « x est à l'intérieur de y , mais x et y ne se chevauchent pas », ce qui peut s'avérer utile pour caractériser les relations de régions concaves ou trouées. Trois nouveaux prédicats asymétriques sont alors ajoutés corres-

pondant aux relations « à l'intérieur », « partiellement à l'intérieur » et « à l'extérieur » :

$$\text{INSIDE}(x,y) \stackrel{\text{def}}{=} \text{DR}(x,y) \wedge \text{P}(x,\text{conv}(y)) \quad (4.6)$$

$$\text{P} - \text{INSIDE}(x,y) \stackrel{\text{def}}{=} \text{DR}(x,y) \wedge \text{PO}(x,\text{conv}(y)) \quad (4.7)$$

$$\text{OUTSIDE}(x,y) \stackrel{\text{def}}{=} \text{DR}(x,\text{conv}(y)) \quad (4.8)$$

L'ajout de cette primitive augmente la taille de l'ensemble des relations possibles, maintenant composé de 23 relations dont 9 correspondent aux combinaisons possibles des relations de convexité et de la relation **EC**, 8 correspondent aux combinaisons avec **DC** et 6 correspondent aux relations restantes échangées de RCC-8. Ce nouveau modèle est nommé RCC-23.

4.2.1.4 Relations de distance et d'orientation

Les relations spatiales présentées ici de façon beaucoup plus succincte sont les relations de distance et d'orientation. Les relations de distance modélisent des expressions telles que « proche de », « loin de ». Ces relations sont souvent modélisées par des approches quantitatives basées sur des mesures de distances de type métrique euclidienne entre entités spatiales. Le problème se pose alors de définir des points d'origine et de destination lorsque les entités spatiales ne sont pas ponctuelles mais surfaciques. Il existe aussi de nombreuses approches hybrides où les distances sont représentées comme des mesures dans des espaces discrétisés et correspondent donc à des degrés de proximité entre entités spatiales. On peut citer les approches issues du traitement d'images et basés sur la morphologie mathématique, comme l'approche granulométrique des travaux de [Mering \(1990\)](#) ou l'approche floue des travaux de [Bloch \(2002\)](#). La discrétisation de l'espace peut être poussé plus loin pour formaliser des distances relatives en raisonnement spatial qualitatif : ici ce sont des relations ternaires qui sont prises en compte pour exprimer des notions de distance via « x est plus proche de y que de z », « x et y sont à la même distance de z » ([Renz et Nebel, 2007](#)).

Les relations d'orientation modélisent, elles, les expressions « à droite de », « à gauche de », « en haut de », « en bas de ». Elles ont pour but de représenter les notions d'alignement, de parallélisme ou d'orthogonalité ([Bonin et Poupeau, 2006](#)). Trois formalismes peuvent être utilisés pour définir les relations d'orientations ([Renz et Nebel, 2007](#)) : 1) les orientations absolues où le système de coordonnées de référence est externe (les directions cardinales) ; 2) les orientations intrinsèques où l'entité spatiale étudiée fixe le repère ; 3) les orientations contextuelles où le repère est fixé par rapport à une entité distincte de l'entité étudiée. Dans la pratique l'étude des relations d'orientation se fait donc en imposant une nouvelle grille de lecture à l'espace contenant les entités spatiales, qui peut prendre la forme d'une grille régulière à mailles carrés ou d'une grille radiante ou d'une grille radioconcentrique. L'orientation d'une entité spatiale est alors donnée de manière absolue ou relative, qualitative ([Renz et Nebel, 2007](#)).

ou quantitative (Colliot, 2003) en repérant la position ou l'orientation de l'entité dans cette grille. Notons qu'en raisonnement spatial qualitatif, l'imposition de cette nouvelle grille de lecture permet de définir dans un même référentiel les relations d'orientation et les mesures de distances relatives données par rapport à une ou plusieurs entités de référence (Stock, 1997).

En reprenant les principes de modélisation par raisonnement spatial qualitatif introduits ici, nous allons maintenant voir comment inférer les modèles chorématiques définis précédemment à partir des images classées des tissus urbains des 50 aires urbaines.

4.2.2 Règles d'identification des morphotypes

Le cadre de modélisation théorique et les outils d'analyse étant posés, il est maintenant possible de définir le cadre de raisonnement spatial qui nous permettra de reconstruire des images classées qui traduisent les effets de combinaison des chorèmes urbains. La démarche proposée consiste à définir les effets de combinaisons des chorèmes par un enrichissement de la sémantique associée aux tissus urbains qui peuvent être interprétés selon la nomenclature de morphotypes introduite précédemment.

4.2.2.1 Notre espace de raisonnement spatial

Avant d'exposer les traitements appliqués aux images classées des tissus urbains, nous introduisons l'espace de représentation et de raisonnement spatial qualitatif que nous utiliserons pour inférer les modèles d'organisations morphologiques.

Représentation des images En reprenant un mode de représentation de l'espace inspiré de celui de la théorie RCC de Cohn et al. (1997), nous définissons différentes entités spatiales composant l'espace des images classées en tissus urbains. Ces images peuvent être considérées comme étant composées des entités suivantes :

- **Espace Physique** : espace correspondant au périmètre des aires urbaines utilisé pour délimiter les classifications.
- **Espace Nul** : espace correspondant à l'espace (ou à une portion d'espace) non bâti, espace complémentaire non labellisé dans nos classifications en tissus urbains.
- **Région** : espace bâti correspondant à un ensemble de pixels adjacents et de même label. Ces entités sont définies en reprenant les propriétés qu'une Région peut avoir dans la théorie RCC, mais en contraignant nos **Régions** à ne pas être composées d'espaces bâtis spatialement distincts : elles sont forcément connexes.
- **Agglomération** : espace bâti correspondant à au moins deux régions qui peuvent

être directement adjacentes ou relier localement par une ou deux mailles contenant des cours d'eau. Ces entités sont définies en reprenant exactement les propriétés qu'une Région peut avoir dans la théorie RCC.

La distinction entre **Région** et **Agglomération** est nécessaire pour raisonner sur les classifications en tissus urbains où l'information apportée par les labels crée des formes supplémentaires. Contrairement à une **Région**, une **Agglomération** peut être composée d'espaces bâtis spatialement distincts puisque dans les classifications, la résolution de 200 mètres du pixel laisse apparaître des ruptures induites par la présence de cours d'eau traversant les agglomérations. Ces déconnexions, bien que significatives en termes d'organisation, ne doivent en effet pas masquer l'unité de ces parties qui ne peuvent être interprétées séparément d'un point de vue urbanistique.

Relations topologiques Le modèle que nous considérerons pour traiter les classifications en tissus urbains et enrichir leur description est une version réduite du modèle RCC-23 présenté précédemment. En effet, nous pouvons éliminer certaines relations qui n'existent pas dans notre cadre de travail. Les objets **Région** étant le résultat d'une partition de l'**Espace Physique** des aires urbaines, il n'existe aucune relation d'inclusion entre des paires de régions : seules les relations **DC** et **EC** peuvent être représentées entre des régions x et y . Les relations sont un peu plus variées entre des régions x et des agglomérations Y , mais certaines relations n'existent pas (les relations **EC** et **PO**) puisque les objets **Agglomération** sont des regroupements de **Régions**. De plus, la distinction entre les relations **INSIDE** et **P – INSIDE** ne sera pas nécessaire dans la pratique, ce qui nous pousse à ne conserver que la première.

Nous nous ramenons ainsi à un modèle décrivant toutes les configurations spatiales en 12 relations, exprimées selon les fonctions booléennes ou opérateurs dilatation $D_B(x)$ dans le tableau 4.2. En plus de ces relations, nous définissons les propriétés supplémentaires suivantes, nécessaires pour caractériser les relations spatiales entre **Régions** en tenant compte de leurs labels :

- **Label**(x) : Label de la région x
- **L.Voisins**(x) : Liste des labels des voisins de la région x
- **L.Voisin+**(x) : Label du voisin de x de plus grande surface
- **Frontiere+**(x) : Label modal sur la frontière extérieur de la région x
- **Mode**(B_{x_i}) : Valeur modale des labels dans le voisinage défini par B_{x_i} centré sur la maille x_i de la région x

Ces propriétés reposent sur l'identification de l'ensemble des pixels **Frontiere**(x) composant la frontière extérieure de la région x et de l'ensemble **Voisins**(x) des régions voisines de x donnés par :

$$\text{Frontiere}(x) = [D_B(x) - x] * I_c \quad \text{où } I_c \text{ est l'image classée} \quad (4.9)$$

$$\text{Voisins}(x) = [D_B(x) - x] * I_n \quad \text{où } I_n \text{ est l'image des régions numérotées} \quad (4.10)$$

TABLE 4.2 – Relations topologiques considérées dans notre représentation

(Région, Région)	
$\text{DC}(x,y)$	$x * D_B(y) == 0$
$\text{EC}(x,y)$	$x * D_B(y) == 1$
(Région, Agglomération)	
$\text{DC}(x,Y)$	$x * Y == 0$
$\text{EQ}(x,Y)$	$x == Y$
$\text{PP}(x,Y) [\text{PPi}(Y,x)]$	$x * Y == x \wedge x! = Y$
$\text{TPP}(x,Y) [\text{TPPi}(Y,x)]$	$\text{PP}(x,Y) \wedge [D_B(x) - x] * [D_B(Y) - Y] == 1$
$\text{NTPP}(x,Y) [\text{NTPPi}(Y,x)]$	$\text{PP}(x,Y) \wedge [D_B(x) - x] * [D_B(Y) - Y] == 0$
$\text{INSIDE}(x,Y) [\text{INSIDEi}(Y,x)]$	$x * Y == 0 \wedge x * \text{conv}(Y) == 1$
$\text{OUTSIDE}(x,Y) [\text{OUTSIDEi}(Y,x)]$	$x * Y == 0 \wedge x * \text{conv}(Y) == 0$

Distance et orientation Dans notre cadre de raisonnement spatial, les relations de distance et d'orientation sont plus secondaires que les relations topologiques. L'utilisation des relations topologiques binaires nous permet de qualifier la distance par le simple proxy « est proche, est loin » repéré par les relations topologiques « est connecté, n'est pas connecté ». Notons toutefois que la définition des agglomérations composées de Régions séparées par des cours d'eau peut être interprétée comme la prise en compte d'une distance floue, renvoyant au concept de *proximity space* (Katsaras, 1979). Dans le chapitre suivant, nous ferons toutefois appel à des relations d'orientation pour qualifier la forme des agglomérations. Ces relations d'orientation serviront à déterminer les directions préférentielles de développement depuis le centre d'une agglomération. Elles seront définies intrinsèquement par l'imposition d'un quadrant centré sur la région centrale de l'agglomération.

Cette représentation des images permet donc de prendre en compte les éléments de modélisation chorématique introduits pour représenter les particularités des aires urbaines. Elle fournit le moyen d'élaborer des règles de raisonnement applicable de façon systématique aux différentes images classées.

4.2.2.2 Simplification des classifications en tissus urbains

Pour articuler le passage des chorèmes urbains aux modèles particuliers observés à partir des données, il est nécessaire de passer par une phase intermédiaire de sim-

plification des classifications en tissus urbains. Ces classifications constituent une représentation de la composition urbaine qui permet de former la base d'un cadre de travail construit à l'échelle de la région d'image. Pour homogénéiser cette base d'étude et s'assurer que ce travail à l'échelle de la région soit aussi pertinent que possible, nous appliquons un processus de généralisation cartographique aux classifications pour les rendre plus stables et lisibles en vue d'être transformées en classifications en morphotypes. Ce processus intervient sur deux niveaux : le niveau pixel, le niveau région.

Niveau Pixel A ce niveau la procédure sert à corriger les erreurs de classification de pixels isolés, les éléments isolés qui constituent une source de bruit, ou même d'incertitude (cf. section 3.3). Trois catégories des classifications des tissus urbains sont concernées : les pixels labellisés centre historique qui lorsqu'ils sont isolés correspondent souvent à la mauvaise classification de grandes emprises industrielles ; les pixels labellisés habitat collectif qui lorsqu'ils sont isolés peuvent correspondre à des mauvaises détections ou à des opérations d'aménagement de moindre ampleur ; les pixels labellisés habitat individuel dispersé qui par définition sont isolés ou définissent des régions de petites tailles et génèrent un bruit important pour la lecture des cartes. L'habitat dispersé ne sera plus représenté dans la suite pour des questions de lisibilité et d'objectif de modélisation. La simplification des classifications est donc opérée à ce niveau en reclassant les deux premières catégories par lissage majoritaire et en supprimant la dernière, selon le processus de la figure 4.9. Le résultat de la transformation est illustré en figure 4.11 sur l'agglomération de Lorient.

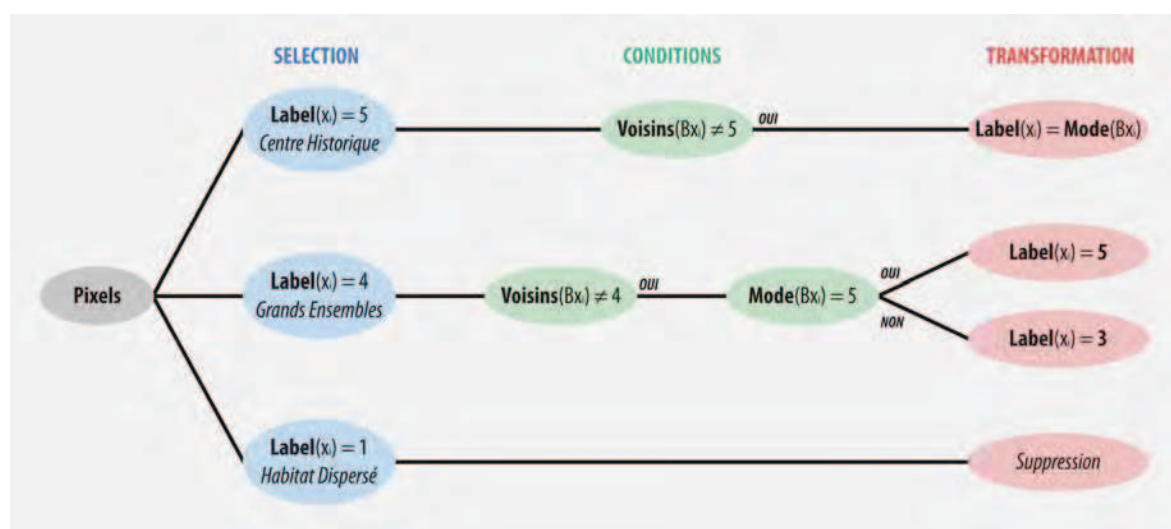


FIGURE 4.9 – Post-traitement de simplification à l'échelle du pixel

Niveau Région A ce niveau la procédure sert à lisser des régions de petite taille qui ne contiennent pas une information significative et structurante, et qui peuvent donc gêner la lecture des nouvelles cartes. Lorsque ces régions sont intégrées à une

agglomération il est préférable de les reclasser pour ne pas créer de trous, mais lorsqu'elles sont isolées il est préférable de les supprimer pour améliorer la lisibilité. Pour conserver la pertinence des classifications, cette procédure n'est appliquée qu'aux régions aux labels les plus fréquents, c'est-à-dire aux régions d'habitat individuel groupé, d'habitat mixte et dense et aux zones d'activités. Nous avons fixé empiriquement un seuil sur la taille des régions à 3 mailles au dessus duquel elles sont considérées comme lisibles et significatives. La simplification est ici aussi réalisée par lissage majoritaire ou par suppression des régions comme illustré en figure 4.10. Ici l'ordre d'exécution des règles de simplification a une importance : les petites zones d'activités doivent être reclassées avant les petites régions d'habitat individuel groupé pour ne pas créer des zones d'activités artefacts de taille significative. Le motif région d'habitat individuel connecté uniquement à une zone d'activité étant fréquent, il semble plus judicieux à ce stade de les reclasser en une plus grande région d'habitat individuel. Le résultat de ce traitement de simplification est aussi illustré en figure 4.11.

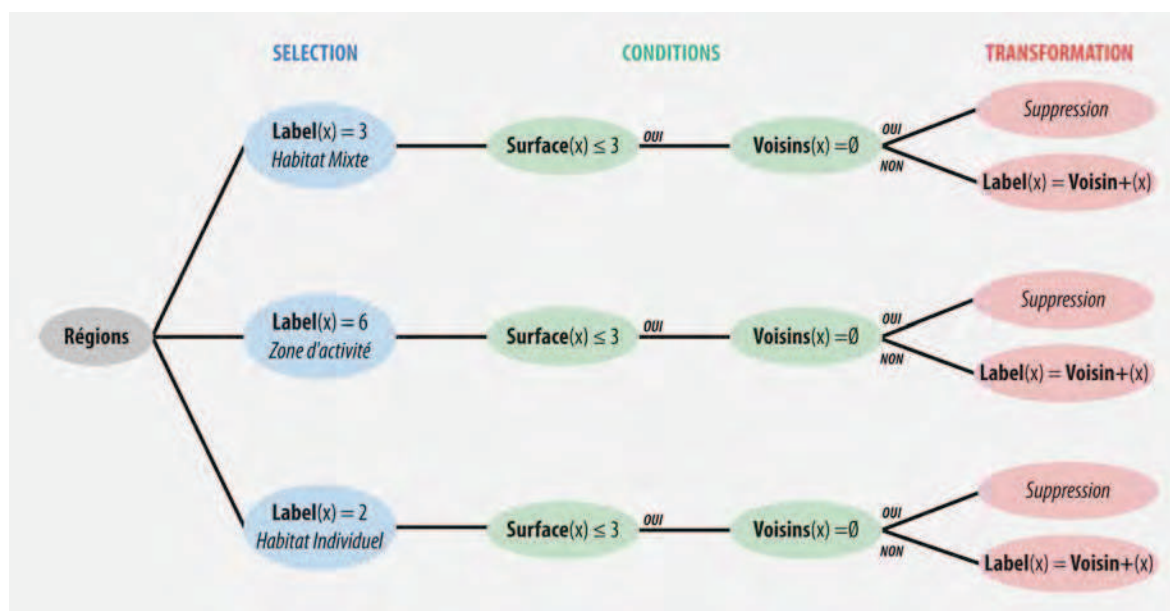
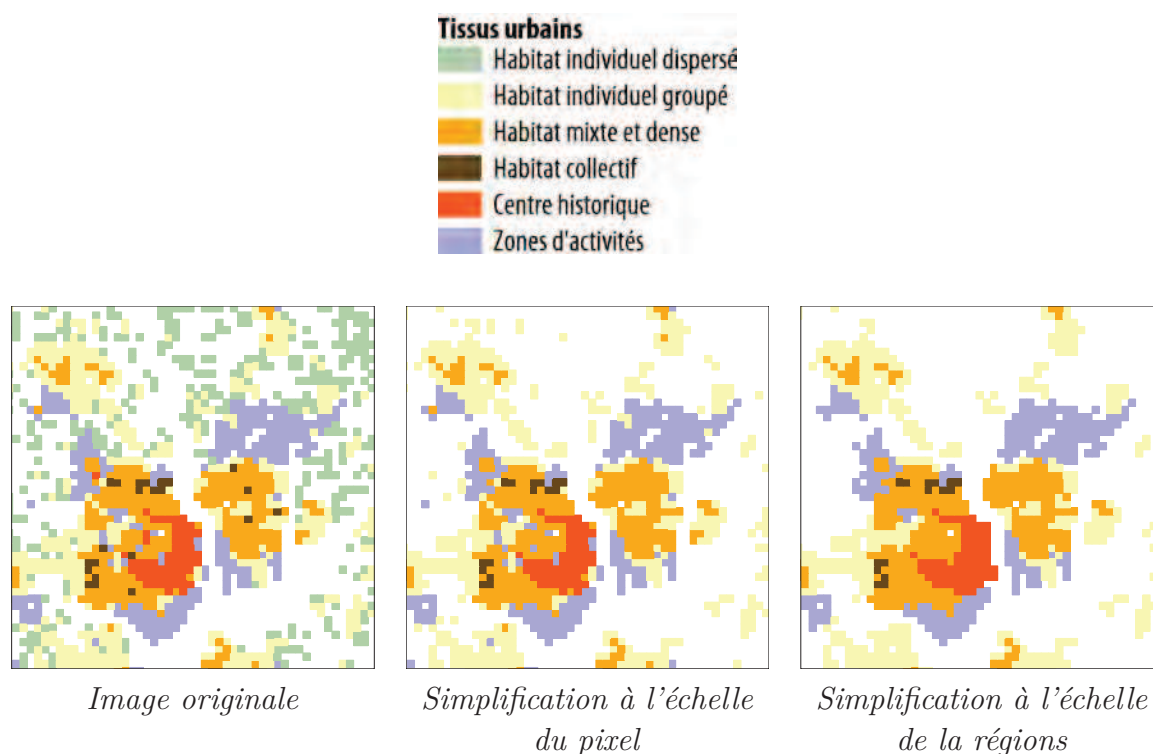


FIGURE 4.10 – Post-traitement de simplification à l'échelle de la région

4.2.2.3 Règles de décision

Sur ces images classées et simplifiées il est maintenant possible d'appliquer des règles de décision pour enrichir les tissus urbains. En utilisant des propriétés de raisonnement spatial il est possible de caractériser plus finement ces régions labellisées pour leur attribuer un morphotype correspondant à la nomenclature associée aux modèles chorématiques de la section précédente. Deux types de propriétés sont utilisées pour caractériser les régions d'image en termes de morphotypes : des caractéristiques de forme des régions, et des caractéristiques relationnelles entre les régions. Les caractéristiques de formes sont utilisées dans un premier temps pour identifier les formes de

FIGURE 4.11 – Étapes de simplification de l'image classée des tissus urbains de Lorient



développement marquées par la coalescence qui est guidée par les réseaux de transport. Les caractéristiques relationnelles sont utilisées ensuite pour re-qualifier chaque région de l'image selon une catégorie de morphotype.

Caractéristiques de forme Certaines agglomérations peuvent être vues comme le fruit de la coalescence de régions à vocation résidentielle anciennement distinctes si l'on raisonne spatialement dans un cadre évolutif. Cette coalescence est observable sur les images classées comme une soudure faible, qui se fait généralement le long d'axes de transport, entre deux morceaux de régions aux surfaces beaucoup plus développées que cette soudure. Nous associons à cette configuration spatiale une indépendance morphologique récemment perdue (compte tenue de l'inertie des espaces urbains) qui matérialise le rôle structurant des réseaux de transport dans le développement des espaces urbains par étalement et intégration. Cette configuration morphologique est intégrée dans la nomenclature de morphotype par la catégorie des noyaux connectés qui ne peuvent pas être identifiés directement sur les classifications.

Les outils du raisonnement spatial qualitatif (Cohn et al., 1997, section 6.4) et de la morphologie mathématique permettent de caractériser ce genre de configuration évolutive, en ordonnant les types de relation qui peuvent être observées entre des paires de régions. Le traitement proposé ici vise à rendre compte de l'état transitoire entre « x et y sont connectés extérieurement » (EC) et « x et y se chevauchent » (PO), et

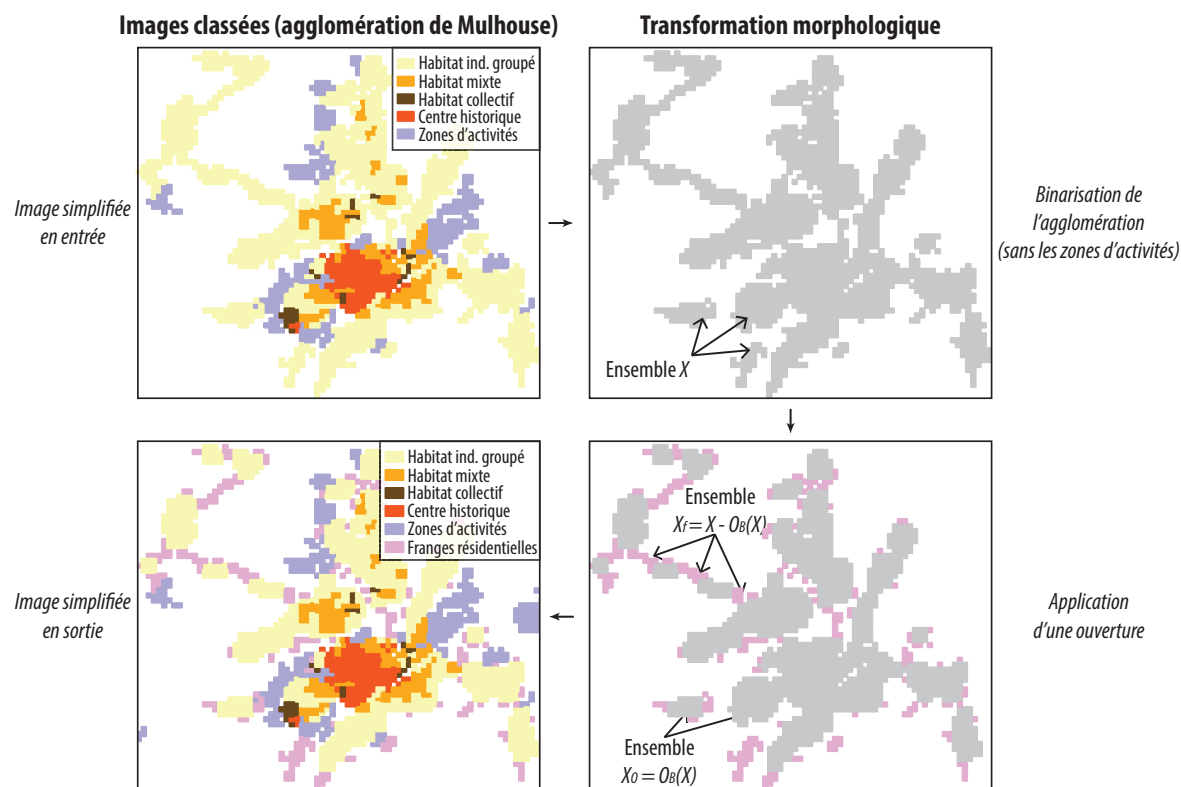


FIGURE 4.12 – Identification des formes d’habitat individuel coalescente sur l’agglomération de Mulhouse

donc d’isoler la zone de chevauchement. Pour cela nous nous intéressons aux régions d’habitat individuel groupé, dernier niveau de densification des espaces urbains, dans lequel nous sommes le plus susceptible d’observer cette coalescence. Nous proposons de la qualifier lorsque ces régions sont intégrées à des agglomérations possédant au moins un centre historique ou une région d’habitat mixte et dense, agglomérations supposées suffisamment développées pour qu’on puisse y déceler des configurations de coalescence significatives². Morphologiquement, nous ne considérons pas les zones d’activités de ces agglomérations, dont la logique d’implantation diffère des zones résidentielles : les zones résidentielles peuvent se développer autour de ces zones qui constituent en quelque sorte un obstacle, toutefois contournable, à leur implantation.

Le traitement, illustré sur l’agglomération de Mulhouse en figure 4.12, consiste donc à décomposer les régions d’habitat individuel groupé en isolant les pixels de l’agglomération correspondant à des zones érodables, c’est-à-dire des soudures de petite taille ou des pixels composant des zones de bordures irrégulières par rapport à l’élément structurant B . En utilisant l’opérateur ouverture O_B sur l’image binaire d’une agglomération X , on obtient :

2. Dans le sens où la coalescence de deux lotissements ne nous semble pas structurellement significative à l’échelle de l’aire urbaine.

mération X (privée de ces zones d'activités), il est possible de répartir les pixels de l'agglomération en deux ensembles :

1. l'ensemble X_O des éléments intacts après l'ouverture $X_O = O_B(X)$,
2. l'ensemble X_f des éléments perdus lors de l'ouverture et récupérer par la transformée $X_f = X - O_B(X)$ ³.

A partir de cette nouvelle partition, il est possible de reclasser temporairement les pixels labellisés habitat individuel groupé selon leur appartenance à l'un des deux ensembles :

1. **Label**(x_i) = 2 si $x_i \in X_O$: les pixels de l'ensemble stable sont laissés inchangés.
2. **Label**(x_i) = 7 si $x_i \in X_f$: les pixels de l'ensemble perdu par ouverture se voient affecter un label temporaire « Frange résidentielle » .

Cette affectation n'est que temporaire puisque comme le montre la figure 4.12, toutes les régions isolées dans l'ensemble X_f n'ont pas le même statut en termes de position et de connexion aux autres régions de l'image classée obtenue en sortie. Certaines jouent un rôle de connecteur entre des régions portant d'autres labels de tissus urbains. Mais d'autres sont en position terminale dans le développement de l'agglomération, elles n'ont pas ce rôle de connecteur. Les caractéristiques de forme permettent donc d'identifier des morphotypes de la nomenclature, mais elles ne sont pas suffisantes pour affecter correctement toutes les régions à des catégories de morphotype. Pour ces régions comme pour toutes les autres régions de l'image, il est systématiquement nécessaire de faire appel à des caractéristiques relationnelles.

Caractéristiques relationnelles Chaque catégorie de tissu urbain apporte déjà des caractéristiques de forme, de fonction et de hiérarchisation. L'ajout des caractéristiques de relations permet d'intégrer la dernière dimension nécessaire pour obtenir une description des structures d'organisations urbaines. La prise en compte de relations entre les régions labellisées est la caractéristique clé, utilisée ici pour définir des règles de décision qui permettent d'identifier chaque catégorie de morphotype des modèles chorématiques.

La figure 4.13 présente l'arbre de décision mis au point pour identifier toutes les catégories de morphotypes à partir des classifications en tissus urbains. Cet arbre permet de transformer totalement la partition en six classes de tissus urbains présents à l'issue du processus de simplification et de caractérisation des formes, en une nouvelle partition en quatorze classes de morphotypes. Les règles de décision qui le constituent doivent être exécutées dans l'ordre de lecture de haut en bas pour assurer que la transition entre les deux images classées s'effectue de manière cohérente et sans créer d'artefacts.

Pour inférer les modèles chorématiques tels que présentés dans la section précédente, des hypothèses ont été posées pour définir la répartition des régions dans chaque ca-

3. Cette transformée est équivalente à la transformée *White Top Hat* définie pour des images en teintes de gris (Serra, 1982).

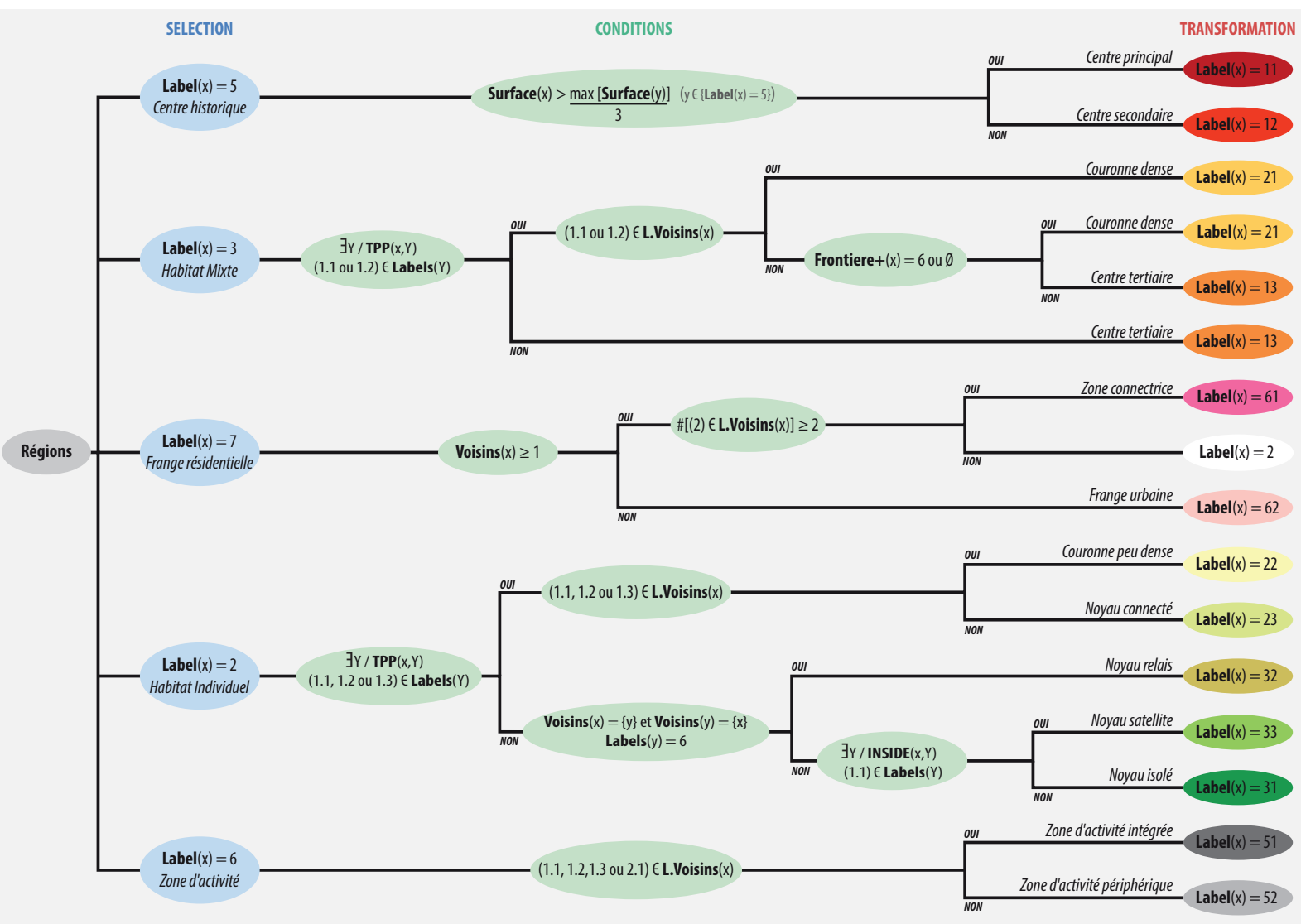


FIGURE 4.13 – Règles de décision pour passer des tissus urbains aux morphotypes

tégorie de morphotypes. Chaque catégorie de tissu urbain est répartie dans différents morphotypes, à l'exception de la catégorie de l'habitat collectif que nous faisons correspondre sans caractéristiques relationnelles au morphotype des grands ensembles.

En ce qui concerne les centres, si une image ne contient qu'une région portant le label centre historique, alors cette région définit le plus haut niveau de la hiérarchie urbaine qualifié de centre principal. Si plusieurs régions portent ce label, alors la hiérarchie est raffinée : les centres principaux sont les régions de type centre historique, de surface supérieure au tiers de la surface du plus grand centre historique de l'aire urbaine ; et les centres secondaires sont les régions de type centre historique restantes. Derrière cette répartition se trouve une volonté de représenter à la fois la hiérarchisation et l'organisation en réseau des lieux dans l'idée d'observer des configurations multipolaires. Le troisième niveau de centralité est défini à partir de la catégorie d'habitat mixte et dense. En considérant la position et les relations de voisinages de ces régions, il apparaît que certaines régions correspondent à des variations plus locales de densité qui peuvent révéler la présence de centres de petite taille et moins denses en population et en bâtiments que les centres historiques.

L'identification des couronnes est réalisée à partir des catégories d'habitat mixte et dense et d'habitat individuel groupé. Cette structuration en couronne de développement qui était un *a priori* d'organisation utilisé pour formaliser la typologie de tissus urbains peut maintenant être explicitée en intégrant des informations sur les relations spatiales. En effet, il a été vu au chapitre 2 que les catégories d'habitat mixte et d'habitat individuel s'organisent selon des profils de densité de population et de surface bâtie décroissants par rapport aux centres historiques des agglomérations. Les couronnes sont donc définies comme les régions de ces catégories adjacentes à un centre ou à une couronne de plus haut niveau de densité.

L'identification des noyaux connectés est rendue possible grâce à la précédente étape de caractérisation des formes. Ils correspondent aux régions d'habitat individuel groupé qui se retrouvent connectées à des couronnes peu denses, ou à d'autres noyaux connectés via des régions portant le label temporaire de franges résidentielles. De plus, deux nouvelles catégories de morphotypes sont créées pour caractériser ces régions : les zones connectrices et les franges urbaines. La première regroupe les régions de franges résidentielles qui connectent au moins deux régions d'habitat individuel. La seconde regroupe les franges résidentielles restantes qui peuvent être qualifiées ainsi puisqu'elles constituent de petits espaces résidentiels isolés en périphérie des agglomérations.

Les zones d'activités, repérées relativement aux autres catégories de morphotypes, sont distinguées assez facilement selon leur positionnement intégré à une agglomération ou périphérique. Les régions connectées à un centre quel que soit son niveau ou à une couronne dense sont considérées comme intégrées, et toutes les régions restantes sont considérées comme périphériques. On ne considère pas que les zones d'activités

uniquement connectées à une couronne peu dense sont intégrées à l'agglomération. Cette position à proximité de zones résidentielles dominées par l'habitat individuel, laisse à penser que le développement plus périphérique de ces zones est plus récent.

L'identification des noyaux satellites est un peu plus délicate car elle nécessite de définir les aires d'influence minimales des agglomérations principales, plus restreintes que le périmètre de l'aire urbaine. Les formes circulaires ou hexagonales étant invalidées par de nombreuses études empiriques, nous faisons le choix d'assimiler l'aire d'influence (morphologique) minimale à l'enveloppe convexe d'une agglomération. Cette enveloppe convexe permettra de définir un espace de proximité qui laisse peu de doute quant à l'existence de liens de subordination forts. Le morphotype des noyaux satellites est alors identifié parmi les régions d'habitat individuel groupé non incluses dans une agglomération contenant des centres, mais incluses dans l'enveloppe convexe d'une agglomération contenant un centre principal.

Les noyaux relais, correspondant à des pôles périphériques associés à des zones d'activité sont identifiés parmi les régions d'habitat individuel groupé non incluses dans une agglomération contenant un centre, mais connectées exclusivement à des zones d'activités. Enfin les noyaux isolés sont définis comme toutes les régions d'habitat individuel groupé qui n'ont pas été affectées à un des morphotypes précédents.

4.2.3 Évaluation des classifications : interpréter les différences

Les règles de décision élaborées de façon à formaliser la nomenclature des morphotypes définissant des modèles chorématiques nous ont permis de construire une représentation graphique des organisations morphologiques des 50 aires urbaines de l'échantillon d'étude. Cette représentation constitue un nouvel atlas de configurations morpho-typiques, donné de façon complète dans l'annexe cartographique (cf. volume II).

Nous ne proposons pas ici d'évaluer les résultats cartographiques de cet atlas en les comparant avec d'autres bases de données puisqu'à notre connaissance il n'en existe pas qui soient suffisamment comparables. Formées à partir de données agrégées au niveau communal, ou désagrégées jusqu'à l'échelle du bâtiment, les classifications les plus comparables pourraient être celles décrivant les espaces urbains selon des logiques de hiérarchisation des lieux et de dynamique d'organisation centre-périphérie. On retrouve dans ce cas le zonage en aire urbaine lui-même (Brutel et Levy, 2011) qui fournit une classification basée sur des critères de hiérarchisation vue par l'emploi. D'autres classifications de communes concernent par exemple une qualification détaillée des zones périurbaines et rurales (Abrantes et al., 2010; Hilal et al., 2012) qui correspondrait aux différents types de noyaux d'urbanisation ou à l'espace complémentaire que nous ne qualifions pas. La classification de Sahraoui et al. (2014) constitue un autre exemple

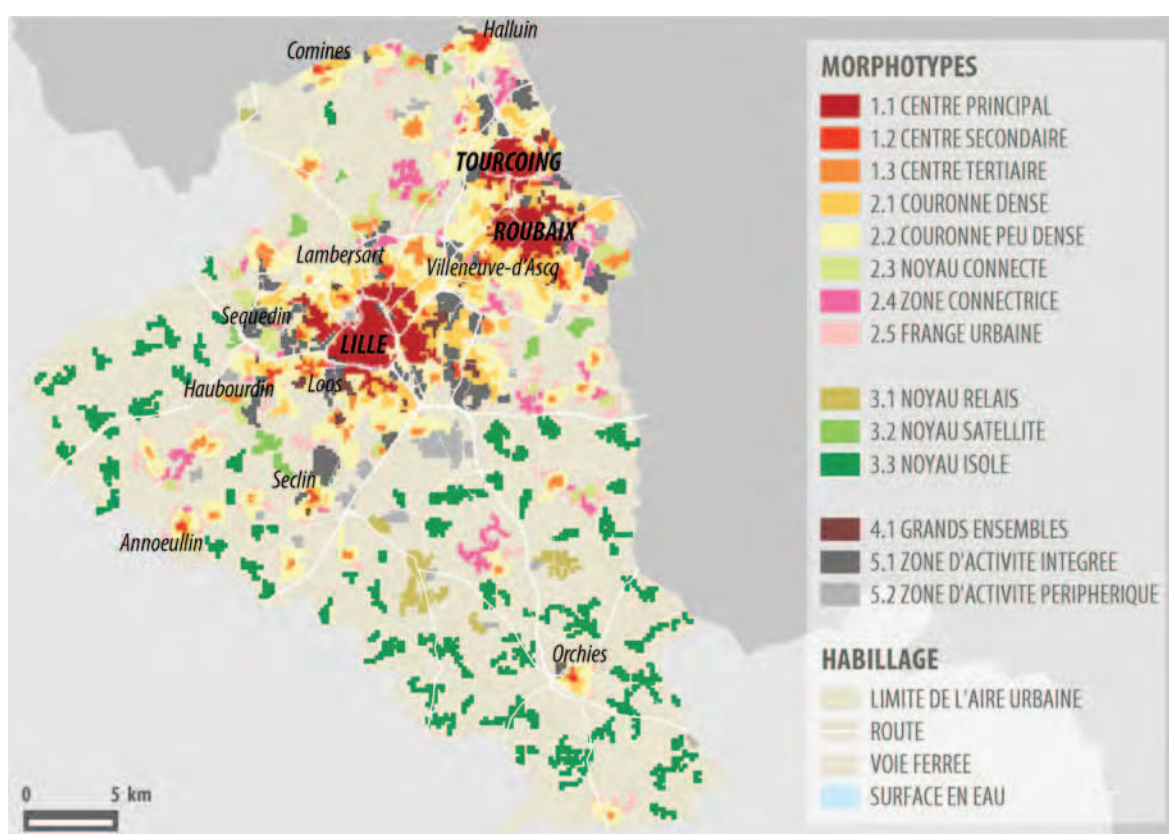
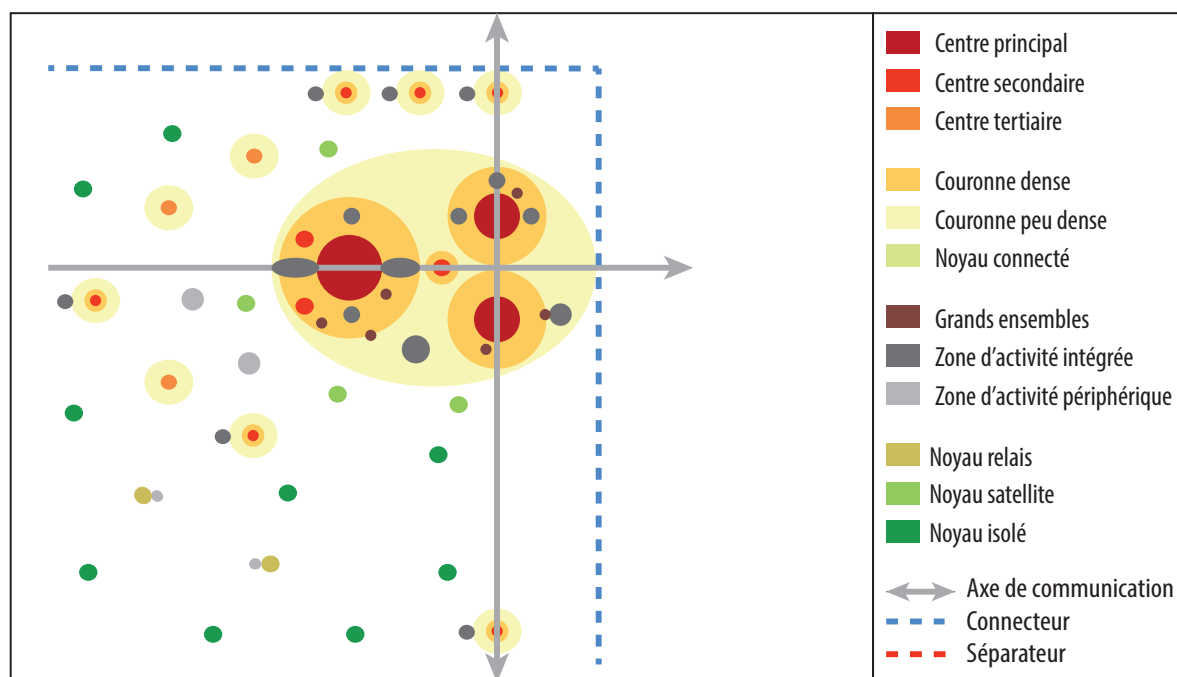


FIGURE 4.14 – Organisation morphologique de l'aire urbaine de Lille (partie française) :
a) Modèle chorématique, b) Classification en morphotypes

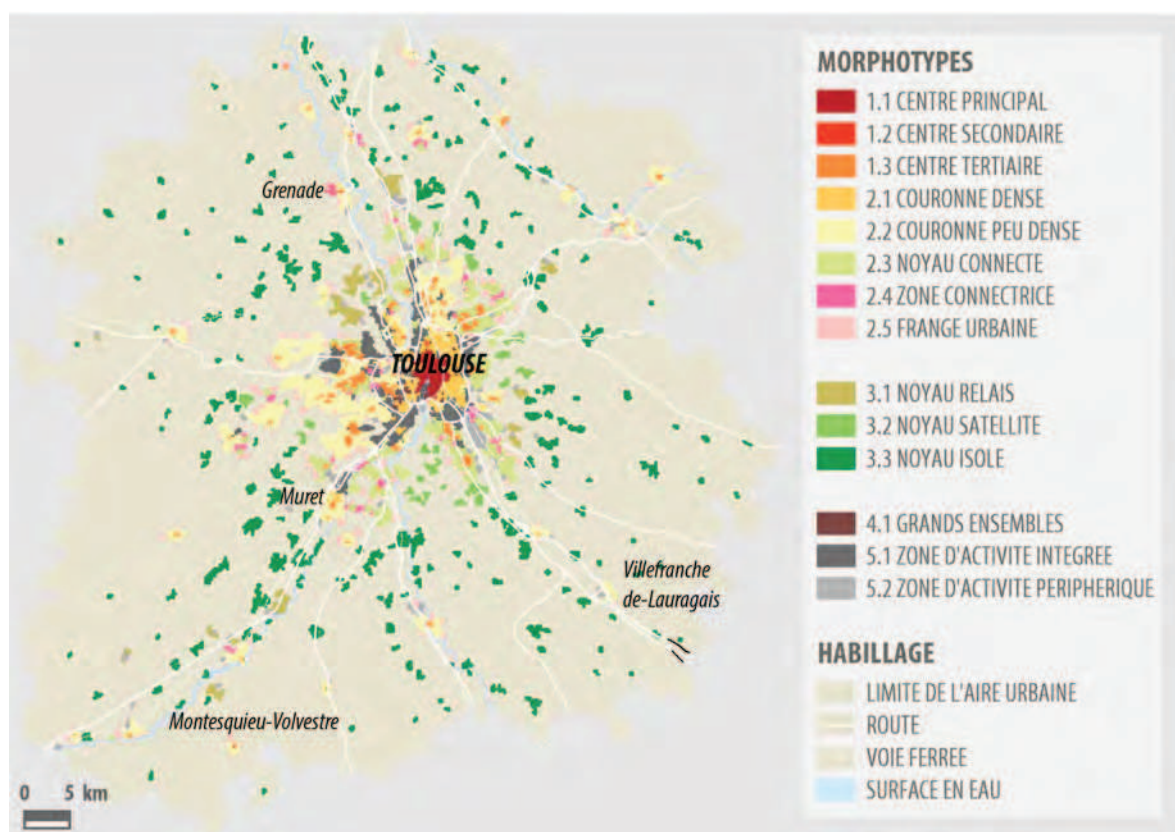
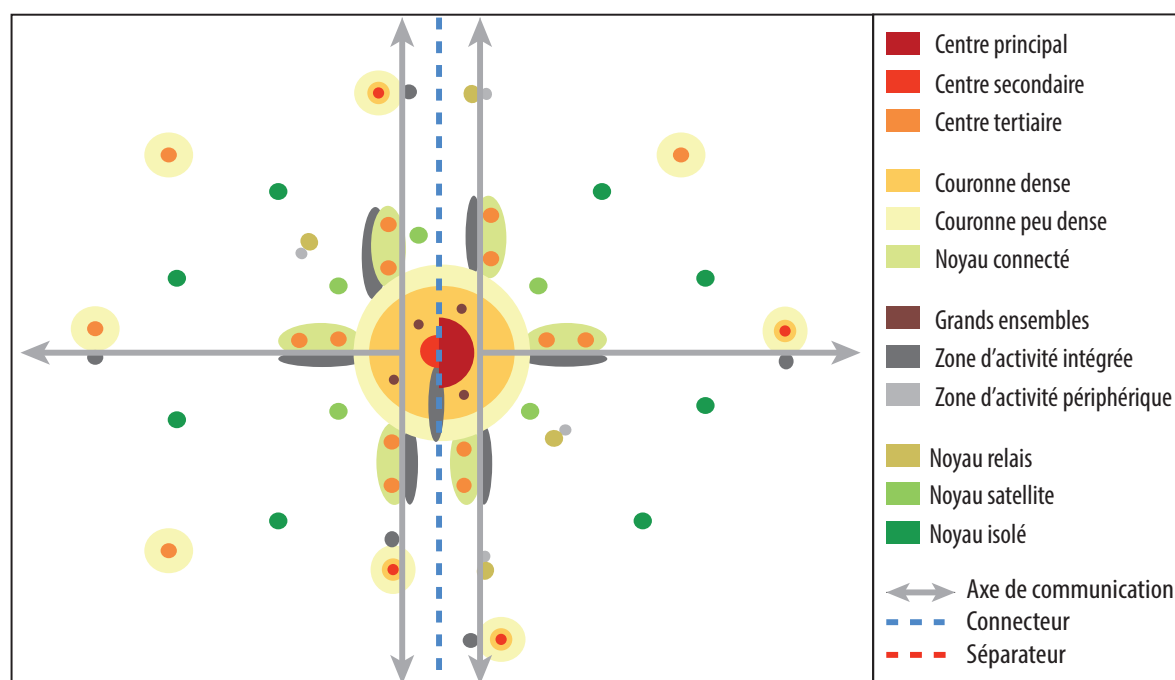


FIGURE 4.15 – Organisation morphologique de l'aire urbaine de Toulouse : a) Modèle chorématique, b) Classification en morphotypes

de typologie mêlant dynamique territoriale et morphologie du bâti résidentiel. Le problème d'une comparaison spatiale est que ces classifications sont constituées à partir de référentiel de données différents. Elles sont en plus données selon des nomenclatures moins détaillées (centre, banlieue ou couronne et périurbain sont les termes qui apparaissent fréquemment), ou qui révèlent des objectifs différents.

Ce que nous souhaitons apporter ici concerne plutôt l'adéquation entre ce qui est observable sur ces classifications en morphotypes et ce que les combinaisons possibles de chorèmes suggèrent quant aux différences de structurations. L'identification des morphotypes nous permet d'exprimer ce qui est commun à l'ensemble des organisations des aires urbaines. La cartographie de ces morphotypes nous permet d'exprimer comment ces types communs peuvent s'organiser différemment spatialement.

En prenant deux exemples caractéristiques, les images classées et modèles chorématiques des aires urbaines de Lille (figure 4.14) et de Toulouse (figure 4.15), nous souhaitons illustrer que ces classifications révèlent bien des processus de structuration des aires urbaines qui peuvent être associés à des organisations différentes. La *sectorisation de l'espace* exprime maintenant la spatialisation des morphotypes. Elle invite à des lectures différentes de l'organisation de ces deux espaces urbains. Chaque catégorie de morphotype est plus ou moins représentée, plus ou moins fragmentée, ce qui peut être interprété comme un impact différent des autres éléments de modélisation contenus dans les chorèmes :

- Le maillage révèle deux *dynamiques territoriales* différentes : l'une à Lille qui se traduit par des espaces relativement denses et compacts sans phénomène important d'extension par étalement et coalescence ; l'autre à Toulouse révèle au contraire un cœur d'agglomération dense, complété par un important volet de développement par coalescence des espaces résidentiels.
- La répartition des morphotypes informe aussi sur les différentes formes de *hiérarchisation* de ces deux systèmes urbains : Lille, prototype français de l'aire urbaine polycentrique contient trois grands centres principaux accompagnés de nombreuses, bien que plus petites, centralités secondaires ou tertiaires ; Toulouse contient un seul haut lieu de centralité et un très grand nombre de centralités tertiaires notamment associées aux noyaux connectés.
- La prise en compte des *contacts et mises à distance* sur ces aires urbaines permet d'observer le caractère aggloméré, ou coalescent de ces deux espaces urbains. Plusieurs agglomérations distinctes et contenant des centres sont observables dans les deux cas. Mais elles sont toujours de taille bien moins importante que l'agglomération principale, en plus d'en être très éloignées.
- Les *réseaux de transport* routiers informent clairement sur la forme globale des agglomérations : les formes compactes de Lille, Roubaix et Tourcoing semblent contenues dans des ceintures dessinées par le réseau, tandis que la forme étalée

de Toulouse semble suivre le développement en étoile matérialisé par le réseau routier.

- Les phénomènes de *ruptures et dissymétries* semblent plus marqué sur l'organisation de Toulouse que de Lille : la Garonne crée un axe de dissymétrie qui explique le dédoublement du Centre historique en un centre principal situé sur la rive droite (terrasse accueillant le développement historique depuis l'époque romaine) et un centre secondaire situé sur la rive gauche (zone longtemps inondable) et le développement plus dense et en continu sur la rive droite que sur la rive gauche.

Du point de vue de la comparaison des organisations, le processus d'interprétation des cartes montre que l'identification des morphotypes ne révèle pas tout à elle seule. Cette nomenclature de formes fournit une base de représentation homogénéisée des éléments composant les structures. Mais les comparaisons des organisations font appel à des caractéristiques de taille, de forme, de relations ou de hiérarchisation qui doivent cette fois être données à l'échelle macro-géographique et globale de l'aire urbaine. Afin de définir un degré de généralisation intermédiaire entre le modèle chorématique général présenté en début de chapitre, et les modèles particuliers dégagés des classifications en morphotypes, il est donc nécessaire d'enrichir la représentation d'une dernière échelle d'étude des organisations morphologiques. Cette représentation graphique sera donc complétée par une description quantitative des classifications qui permettra d'intégrer et de résumer l'interprétation et les comparaisons qui peuvent se faire à partir de ces cartes.

Conclusion

Ce chapitre nous a permis de construire une représentation des structures urbaines, basée sur les précédents travaux de classification des tissus urbains des 50 aires urbaines de l'échantillon. En ajoutant une sémantique relative aux relations spatiales entre les régions labellisées de ces classifications il a été possible de construire une représentation graphique et cartographique, de ce que peuvent être les structures de ces aires urbaines observées d'un point de vue morphologique et topologique.

Cette représentation est le fruit d'un dialogue entre modèles théoriques formalisés sous forme de chorèmes et modèles expérimentaux constitués à partir des classifications en tissus urbains. Elle a pu être formalisée de façon systématique grâce au raisonnement spatial qualitatif qui a permis d'inférer le processus de représentation des modèles chorématiques et d'exploration des données. Ces aires urbaines sont donc maintenant représentées à l'échelle de la région d'image selon une nomenclature de 14 morphotypes décrivant les formes de développement urbain.

Cette nouvelle représentation constitue une nouvelle base d'étude des organisations urbaines qui peut être exploitée pour spatialiser différentes thématiques de données relatives aux aires urbaines. L'objectif du chapitre suivant sera donc d'exploiter ces représentations à l'échelle de l'image pour identifier les phénomènes d'organisation les plus pertinents et discriminants spatialement parmi les 50 aires étudiées. L'idée est maintenant de tester la significativité de ce maillage morphologique pour travailler à l'échelle de la forme urbaine métropolitaine et construire une typologie des différentes organisations.

Chapitre 5

Typologie des structures urbaines

Sommaire

Introduction	196
5.1 Une représentation quantitative des structures urbaines	196
5.2 Méthodes d'analyse factorielle	205
5.2.1 Les analyses factorielles multiples	206
5.2.2 Extensions de l'AFM aux données mixtes	211
5.3 Identification des tendances spatialement structurantes	215
5.3.1 Les analyses partielles des groupes de variables	216
5.3.2 L'analyse globale des caractéristiques d'organisation	226
5.3.3 Construction de la typologie par CAH	230
5.4 Chorotypes : la spatialisation des enjeux d'aménagement	236
5.4.1 Les modèles en chorotypes	236
5.4.2 Modèles urbains et enjeux d'aménagement	239
Conclusion	243

Introduction

L'objectif de ce chapitre est de chercher à exploiter les précédents travaux de classification de l'organisation morphologique des aires urbaines, dans le but d'identifier différents types de structure urbaine. L'idée est de décrire ces aires urbaines, cette fois-ci étudiées à l'échelle de la macro-forme observée sur chaque image, dans le but de déterminer les caractéristiques marquantes qui différencient les structures. L'objectif final est d'identifier des formes d'organisation qui peuvent être qualifiées de chorotypes, c'est-à-dire des « arrangements récurrents de chorèmes », qui puissent être évalués en fonction des enjeux de gestion et de développement durable pris en compte actuellement dans les politiques d'aménagement urbain.

Pour cette dernière phase de classification, les images classées en morphotypes nous serviront de base de travail. Nous associerons différents types de mesures quantitatives à chaque catégorie de morphotype dans le but d'obtenir une représentation synthétisant différentes caractéristiques morphologiques. En travaillant sur une représentation mêlant mesures qualitatives et quantitatives, nous faisons le choix d'explicitier par la partition de l'espace et les étiquettes qualitatives associées, les *a priori* qui viendront guider l'interprétation des structures d'organisation urbaine. Les données quantitatives serviront, elles, à mesurer la variété des organisations et à les comparer en produisant une typologie.

Nous verrons donc en introduction de ce chapitre les mesures retenues pour quantifier différents volets de l'organisation des aires urbaines. Après un bref état de l'art des méthodes d'analyse factorielle multiple, nous présenterons la méthode d'analyse factorielle de données mixtes utilisée pour traiter les données et extraire des mesures résumant les tendances caractérisant l'organisation des 50 aires urbaines étudiées. Nous présenterons ensuite les résultats de l'analyse factorielle et de la CAH appliquée pour identifier différents groupes de structures urbaines, ou chorotypes. Enfin nous concluons ce chapitre et ce travail en mettant en perspective les résultats et les chorotypes obtenus par rapport à différents enjeux d'aménagement identifiés dans différents documents de planification urbaine.

5.1 Une représentation quantitative des structures urbaines














A l'issue du chapitre précédent, nous avons obtenu un ensemble d'images classées en morphotypes qui révèlent des différences significatives d'organisation parmi les 50 aires urbaines de l'échantillon d'étude. Ce travail précédent nous a permis d'examiner de manière intuitive les possibilités de représentation de différents modèles théoriques décrivant les structures urbaines à partir de données sur grille régulière. Ces modèles

théoriques confrontés aux configurations urbaines observables révèlent une diversité de structurations qui peuvent impacter aussi bien le développement urbain que les pratiques de déplacement. Cette variété pour l'instant remarquée par interprétation et comparaison des cartes doit être testée de manière plus systématique afin de déterminer une typologie reprenant les caractéristiques les plus marquantes de ces 50 structures urbaines.

Nous introduisons ici un espace de caractéristiques mêlant tableaux de fréquence et mesures quantitatives pour améliorer la caractérisation des structures urbaines et donc définir une typologie. L'utilisation de ces mesures nous permet de proposer un cadre de travail facilitant la généralisation et la comparaison simultanée des 50 aires urbaines étudiées. Dans le contexte d'un travail présenté comme un va-et-vient entre modélisation théorique et expérimentale, la constitution de vecteurs de caractéristiques basés sur des mesures de nature mixte correspond à une volonté d'intégrer les résultats de la précédente phase cartographique et qualitative, à une phase de test et de différenciation des structures observées (Creswell et Clark, 2010, Chapitre 1).

Les mesures quantitatives présentées ici sont utilisées pour intégrer l'implicite contenu dans l'interprétation des cartes en morphotypes. Le choix de ces mesures est une étape aussi importante que l'étape précédente d'identification des morphotypes puisqu'il orientera la vision que nous nous faisons des structures urbaines, et déterminera donc entièrement la typologie obtenue en sortie. Nous proposons donc d'introduire des mesures résumant des caractéristiques liées à la densité, à la fragmentation ou à la répartition spatiale des morphotypes, des caractéristiques liées à la complexité des formes prises à l'échelle supérieure de l'agglomération, toutes supposées discriminantes pour comparer les organisations.

Les mesures retenues sont donc toutes basées sur la nomenclature en morphotypes (des 50 aires urbaines). Nous reprenons la composition morphologique exprimée à l'aide de 13 catégories de morphotypes permettant de couvrir l'ensemble des espaces bâtis :

	Centre principal		Noyau satellite
	Centre secondaire		Noyau isolé
	Centre tertiaire		Grands ensembles
	Couronne dense		Zone d'activité intégrée
	Couronne peu dense		Zone d'activité périphérique
	Noyau connecté		Habitat individuel dispersé
	Noyau relais		

Deux changements sont notables dans la nomenclature considérée ici : l'habitat individuel dispersé a été réintroduit ; les franges urbaines et zones connectrices ont été supprimées. L'habitat individuel dispersé, qui était difficile à conserver dans une repré-

sentation graphique basée sur l'étude de région, est ré-introduit pour tenir compte d'un niveau de structuration supplémentaire de l'espace complémentaire, qui peut accueillir dans certain cas une part non négligeable de population. A l'opposé, les catégories de franges urbaines et zones connectrices ont été supprimées pour leur caractère plus négligeable en termes d'emprise spatiale et de population.

Se baser sur une discrétisation de l'espace nous permettra de conserver l'apport de la modélisation théorique pour interpréter les différences d'organisation les plus marquantes. La richesse de cette nomenclature morphologique permettra en effet de quantifier les différences structurant les organisations spatiales en fonction des différentes sectorisations des espaces urbains, des différentes dynamiques territoriales, mais aussi des différentes formes de hiérarchisation des territoires.

La répartition de la population

Comme expliqué tout au long de ce travail, la densité est un champ important pour caractériser les structures urbaines aussi bien pour des questions de consommation d'es-

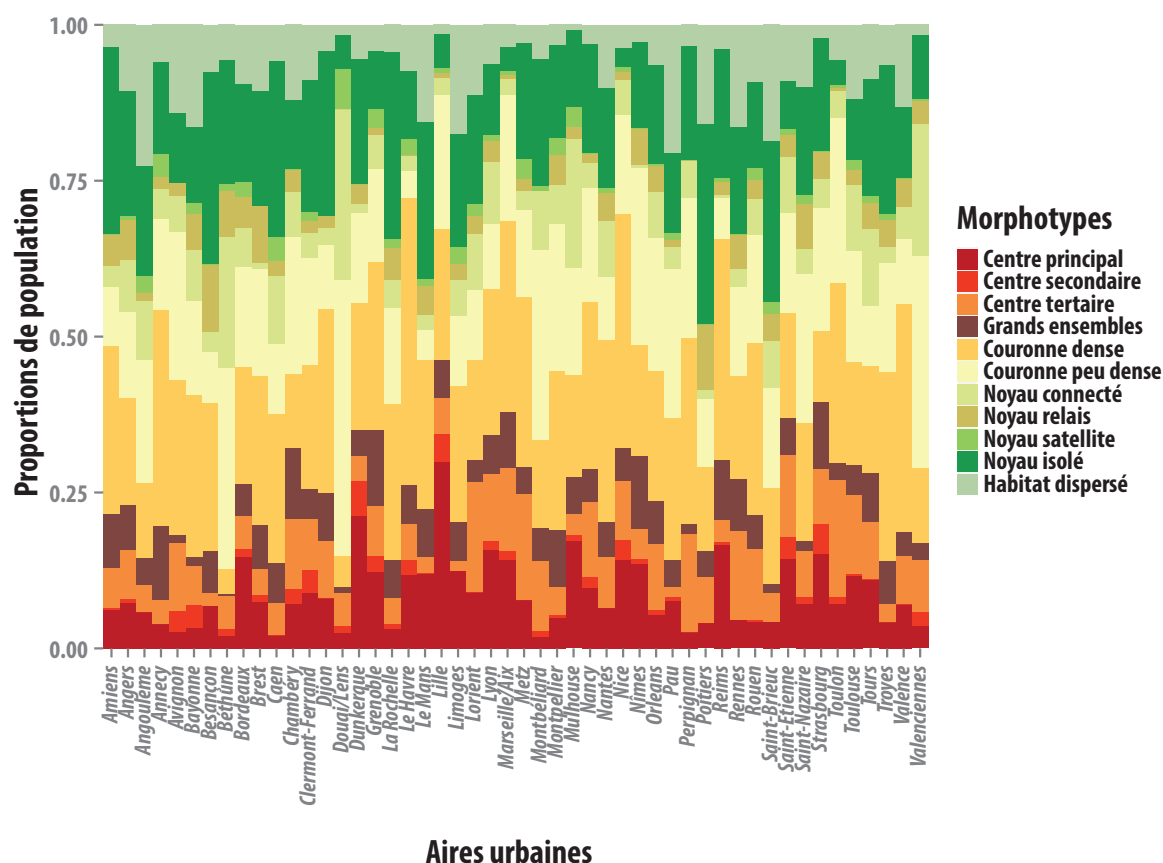


FIGURE 5.1 – Proportions de population par morphotypes résidentiels pour les 50 aires urbaines de l'échantillon

pace, que pour des questions d'exploitation et de valorisation des réseaux de transport. La première variable retenue concerne donc la répartition de la population au sein des aires urbaines. Elle est donnée pour un morphotype i par rapport à l'ensemble \mathcal{R} des régions x d'une image classée :

$$P_i = \frac{\sum_{x \in \mathcal{R}_i} \mathbf{Population}(x)}{\sum_{y \in \mathcal{R}} \mathbf{Population}(y)} \quad \text{où } \mathcal{R}_i = \{x \in \mathcal{R} | \mathbf{Label}(x) = i\} \quad (5.1)$$

Sachant que les deux catégories de zones d'activités ne seront pas considérées puisqu'elles sont très peu, ou pas habitées et qu'elles ne constitueraient donc pas des modalités discriminantes en termes de répartition de la population.

Cette variable de population qui a servi à la construction des classifications en tissus urbains est implicitement contenue dans les morphotypes. Mais l'objectif en la ré-introduisant est d'observer le poids relatif de chaque morphotype par rapport à l'ensemble de la population d'une aire urbaine. De plus, bien que le facteur de taille des territoires urbains constitue une source importante de différenciation (Schwarz, 2010), il semble important de s'en affranchir pour assurer la comparabilité des aires urbaines de l'échantillon. Le système des aires urbaines françaises étant très fortement hiérarchisé (cf. Annexe, volume II), il semble plus judicieux de s'affranchir de ce facteur de taille pour ne pas isoler les premières aires urbaines de l'échantillon comme Lyon, Marseille ou Lille de taille 13 fois plus importante que les dernières aires urbaines de l'échantillon. Normaliser la population des régions de morphotypes i par la population de toutes les régions de l'aire urbaine, permet donc de comparer la répartition de la population par morphotypes sur l'ensemble des aires urbaines (figure 5.1).

La fragmentation morphologique des aires urbaines

La deuxième variable retenue concerne la fragmentation des régions associées à un morphotype donné, qui peut être vue comme un critère d'homogénéité des régions. La mesure choisie utilise un critère de comptage des régions ordonnées par leur surface pour quantifier la fragmentation des régions les plus importantes en surface. Les régions de plus petites tailles sont négligées pour limiter la prise en compte des effets de morcellement qui peuvent être attribués aux résultats de la segmentation d'image. En considérant les x_k régions ordonnées ($\mathbf{Surf}(x_k) \geq \mathbf{Surf}(x_{k+1})$), la mesure est donnée pour un morphotype i par :

$$S_i = \frac{\#\{x_k \in \mathcal{R}_i | \sum_{x_1, \dots, x_k} \mathbf{Surf}(x_k) = 0,75 \mathbf{Surf}(\mathcal{R}_i)\}}{\#\mathcal{R}} \quad (5.2)$$

Sachant que le morphotype de l'habitat individuel dispersé n'est pas considéré compte tenu de sa nature fragmentée, et que les mesures sont données relativement au nombre total de régions d'une classification pour s'affranchir du facteur de taille des aires

urbaines.

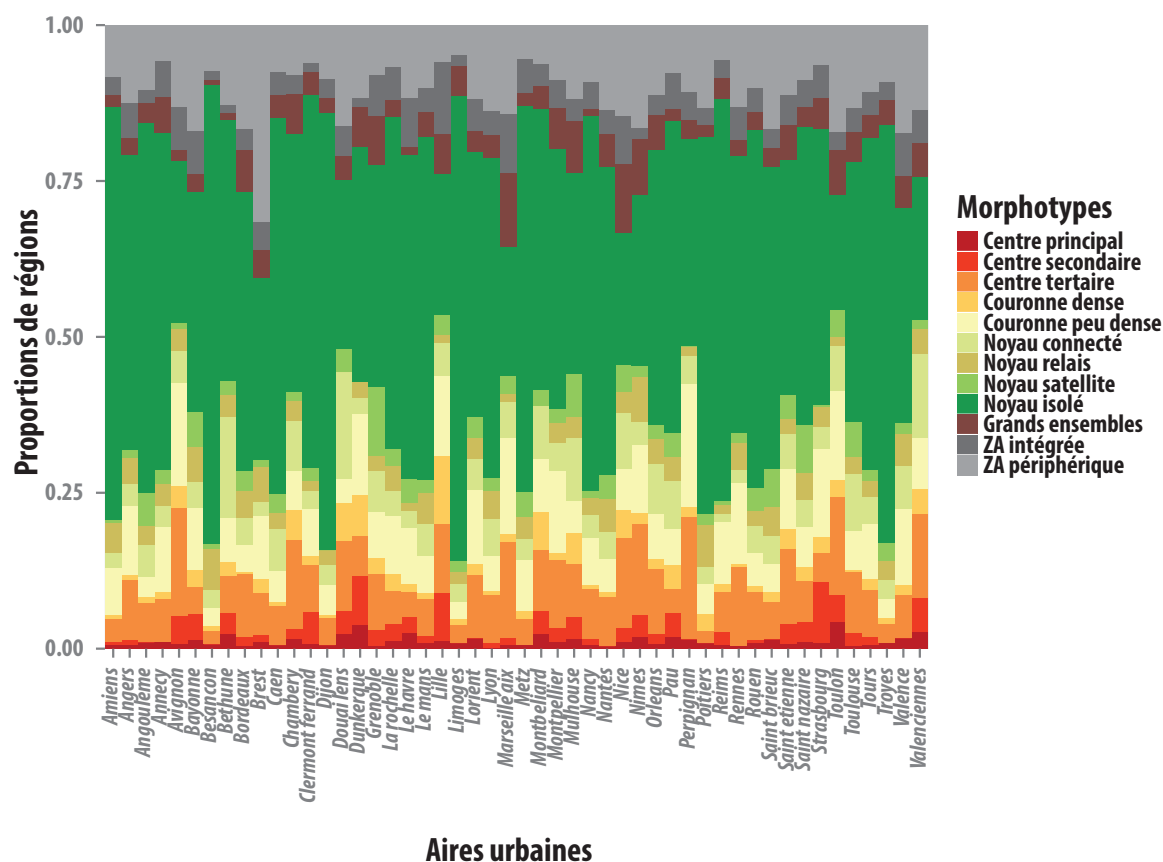


FIGURE 5.2 – Proportions de la fragmentation des régions par morphotypes de chaque aire urbaine de l'échantillon

Cette variable renseigne donc sur les différences d'organisation liées aux coupures présentes dans l'espace urbain. Elles sont notamment liées aux axes de rupture et de dissymétrie structurant les agglomérations et aux effets d'intégration de multiples lieux morphologiquement distincts au sein des formes métropolitaines. Ces coupures engendrées par des phénomènes aussi différents que les contraintes de site, les infrastructures de transport ou l'histoire urbaine informent donc sur la complexité morphologique des aires urbaines (figure 5.2).

Les relations d'adjacence

En complément de ces variables, il semble important de quantifier les relations spatiales entre morphotypes, ainsi que leur positionnement dans le système urbain. L'objectif est d'interroger les différences d'organisation en fonction du rôle de la distance et des contacts afin de quantifier la composition des agglomérations dont la morphologie interne peut être plus ou moins complexe. L'idée est ici de quantifier la dynamique de

développement des agglomérations au travers du positionnement des couronnes et de leur densité, ou bien encore au travers du positionnement des noyaux connectés, ou celui des grands ensembles.

Pour mieux caractériser les organisations, une notion de positionnement dans le système urbain est intégrée pour étudier les différentes agglomérations présentes au sein des aires urbaines. Afin de ne pas produire de trop nombreuses mesures, nous faisons le choix de diviser l'image classée de chaque aire urbaine en deux espaces :

- **Espace central** : ensemble des régions \mathcal{R}_C de l'agglomération accueillant le centre principal de plus grande surface.
- **Espace périphérique** : ensemble des régions \mathcal{R}_P de l'espace complémentaire à l'espace central.

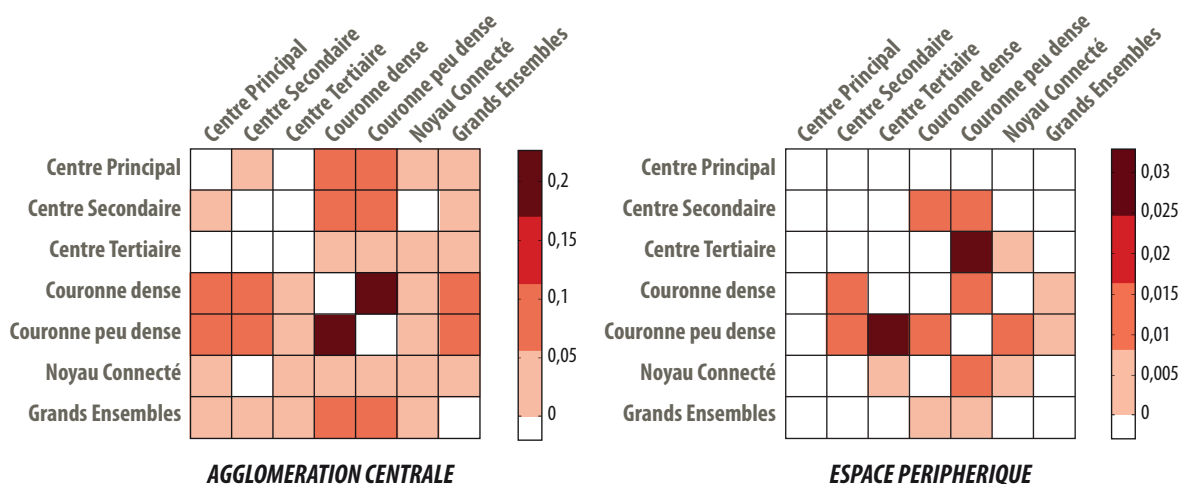


FIGURE 5.3 – Matrices de relations spatiales dans l'aire urbaine de Lille

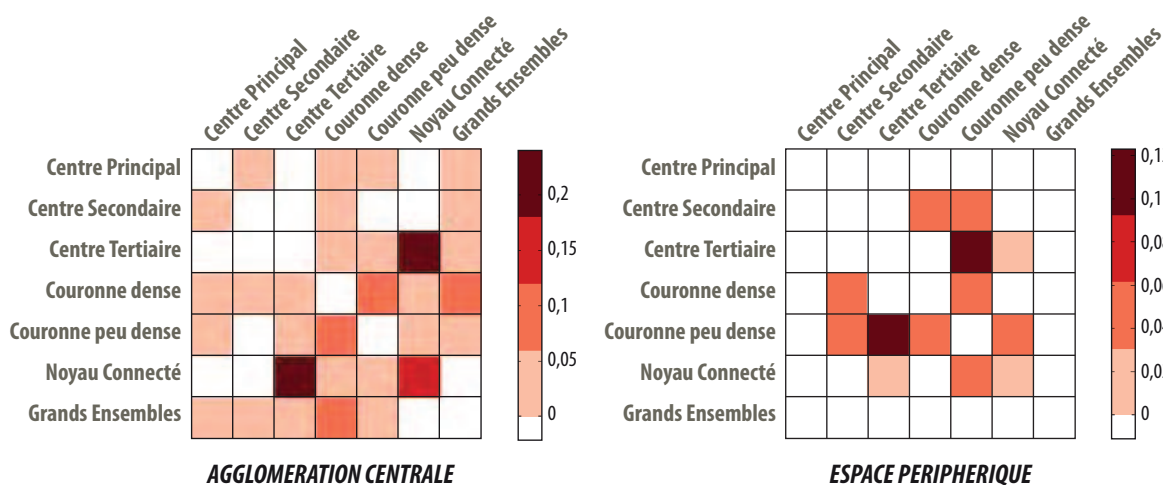


FIGURE 5.4 – Matrices de relations spatiales dans l'aire urbaine de Toulouse

A partir de là, nous nous intéressons aux relations de connexion $\mathbf{EC}(x,y)$, c'est-à-dire les relations d'adjacence entre des paires de régions (x,y) de morphotypes (i,j) appartenant soit à \mathcal{R}_C soit à \mathcal{R}_P . Chaque image classée est caractérisée par deux matrices M_C et M_P qui peuvent être vues comme des matrices d'adjacences pondérées (Gross et Yellen, 2004) ou des matrices de cooccurrence (Sonka et al., 2008) données par :

$$M(i,j) = \frac{\#\{(x,y) \mid x \in \mathcal{R}_i, y \in \mathcal{R}_j, \mathbf{EC}(x,y)\}}{\#\{(x,y) \mid \forall k,l; x \in \mathcal{R}_k, y \in \mathcal{R}_l, \mathbf{EC}(x,y)\}} \quad (5.3)$$

La mesure des relations spatiales dans ces deux espaces complémentaires nous permet donc de composer un système d'agglomérations et d'étudier leurs structurations internes. Pour composer les agglomérations, nous nous basons sur les 7 morphotypes suivants : centre principal, centre secondaire, centre tertiaire, couronne dense, couronne peu dense, noyau connecté, et grands ensembles. Les autres morphotypes ne sont pas considérés pour plusieurs raisons : les autres catégories de noyaux correspondent par définition à des régions morphologiquement isolées ; les deux types de zone d'activité sont déjà définis en fonction de leur connexion à certains morphotypes ; et enfin l'habitat individuel dispersé comme pour les caractéristiques de fragmentation morphologique est moins adapté à ce genre d'étude.

Les matrices des relations spatiales obtenues sont illustrées pour l'aire urbaine de Lille en figure 5.3 et l'aire urbaine de Toulouse en figure 5.4. La grande dimensionnalité de l'espace engendré par ces relations peut s'avérer problématique puisque combinées aux précédentes variables, elles donneraient un espace de dimension plus grande que le nombre total d'aires urbaines. D'autre part, cet espace de caractéristiques peut être source de problèmes lors d'une analyse à visée discriminative si certains types de relations apparaissent trop rarement. Leur rareté globale sur l'échantillon risque de biaiser les résultats en prenant une place trop importante dans des analyses basées sur des critères de variances. Certaines relations telles que les relations entre centres principaux

TABLE 5.1 – Labels des relations d'adjacence étudiées

CENTRE	C11/21	{ Centre principal Couronne dense	C21/22	{ Couronne dense Couronne peu dense	PÉRIPHÉRIE	P12/21	{ Centre secondaire Couronne dense
	C11/22	{ Centre principal Couronne peu dense	C21/41	{ Couronne dense Grands ensembles		P12/22	{ Centre secondaire Couronne peu dense
	C11/41	{ Centre principal Grands ensembles	C22/41	{ Couronne peu dense Grands ensembles		P13/22	{ Centre tertiaire Couronne peu dense
	C12/21	{ Centre secondaire Couronne dense	C23/22	{ Noyau connecté Couronne peu dense		P13/23	{ Centre tertiaire Noyau connecté
	C12/22	{ Centre secondaire Couronne peu dense	C23/23	{ Noyau connecté Noyau connecté		P21/22	{ Couronne dense Couronne peu dense
	C13/22	{ Centre tertiaire Couronne peu dense	C23/41	{ Noyau connecté Grands ensembles			
	C13/23	{ Centre tertiaire Noyau connecté					

et autres morphotypes en périphérie, permettraient d'identifier des structures polycentriques mais la rareté de ces relations qui ne sont présentes en faible nombre que dans quatre aires urbaines (Douai-Lens, Béthune, Valenciennes et Toulon) pousse toutefois à les supprimer. Une sélection de 18 relations discriminantes et suffisamment présentes est retenue : 13 en zone centrale et 5 en zone périphérique (tableau 5.1).

La morphologie des agglomérations

Le dernier thème intégré concerne la forme des agglomérations principales. L'idée est ici d'intégrer des mesures qui permettraient de caractériser l'impact des logiques de développement dissymétrique de la forme, et donc de retrouver la présence ou l'absence de directions préférentielles de développement qui peuvent par exemple influencer les performances des réseaux de transport.

Pour différencier les formes des agglomérations, nous considérons un premier indice de forme, l'excentricité (Sonka et al., 2008), pour évaluer la forme plus ou moins linéaire ou circulaire des agglomérations principales. Pour une région ou agglomération, cet indice correspond à la mesure de l'excentricité de l'ellipse de même moment d'ordre deux que la région ou l'agglomération. L'excentricité de cette ellipse est donnée selon la formule :

$$e = \frac{\sqrt{a^2 - b^2}}{a} \quad \text{où} \quad \begin{array}{l} a = \text{longueur du demi grand axe} \\ b = \text{longueur du demi petit axe} \end{array} \quad (5.4)$$

Prenant ses valeurs entre 0 et 1, l'excentricité mesure l'écrasement d'une ellipse sachant qu'une valeur à 0 indique une forme circulaire tandis qu'une valeur à 1 indique une forme linéaire.

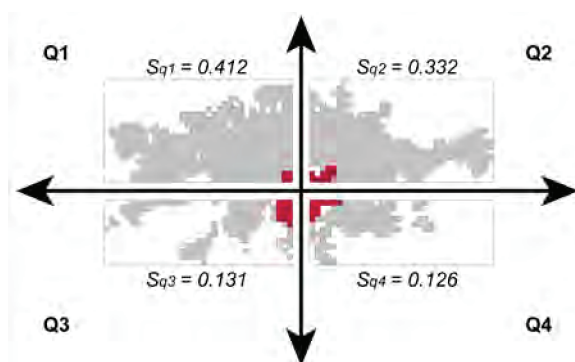


FIGURE 5.5 – Evaluation de la forme de l'agglomération de Besançon

En plus de cet indice global de forme qui peut vite montrer ses limites en cas de forme concave, nous construisons un indice de forme basé sur le découpage de l'agglomération principale en quadrants repérés par rapport au centre principal. L'idée derrière la mise au point de cet indice est de pouvoir évaluer la forme des couronnes

de développement et de discriminer les agglomérations au développement circulaire et non contraint, des agglomérations au développement contraint en demi-couronne par exemple. La figure 5.5 illustre le résultat du positionnement d'un tel quadrant sur le centroïde de la région correspondant au centre principal de l'agglomération de Besançon. Pour obtenir un vecteur de mesures du développement dans chaque partie, nous appliquons le processus suivant aux images binaires des agglomérations principales de chaque aire urbaine :

1. Une rotation est appliquée à l'agglomération de façon à ce que le grand axe de l'ellipse de même moment d'ordre deux que l'agglomération coïncide avec l'axe horizontal de l'image.
2. L'agglomération est découpée en quadrants q_i à partir d'un repère placé sur le centroïde de la région labellisée en centre principal.
3. Les ratios de surface bâtie S_{q_i} sont calculés dans chaque partie du quadrant :

$$S_{q_i} = \frac{\text{Surface}(q_i)}{\text{Surface}(Y)} \quad (5.5)$$

4. Un vecteur SC_q de dimension 3 est formé à partir du cumul des ratios de surface S_{q_i} :

$$SC_q = (S_{q_1}; S_{q_1} + S_{q_2}; S_{q_1} + S_{q_2} + S_{q_3}) \quad \text{où} \quad \begin{matrix} S_{q_1} \geq S_{q_2} \geq S_{q_3} \\ \sum_{i=1}^4 S_{q_i} = 1 \end{matrix} \quad (5.6)$$

Ce vecteur permet de caractériser les directions de développement indépendamment de la taille de l'agglomération et de son orientation globale. Le fait d'introduire une somme cumulée des ratios de surface introduit une corrélation qui nous sert à caractériser globalement le développement de la couronne dans les quatre directions. Les ratios plus élevés que 0,25 pour le premier élément ; 0.50 pour le deuxième ; et 0.75 pour le troisième indiquent une contrainte dans une ou plusieurs directions de développement (1/4 de couronne, demi-couronne ou 3/4 de couronne). A l'opposé, lorsque le vecteur tend à se rapprocher de (0.25; 0.50; 0.75) le développement peut être qualifié d'isotrope ou de couronne entière.

En se basant sur la sectorisation des aires urbaines selon la nomenclature en morphotypes, nous avons déterminé quatre thématiques de caractéristiques numériques qui fournissent les informations nécessaires pour qualifier la morphologie des structures urbaines. Ainsi en prenant en compte sous la forme de mesures mixtes les grands volets d'interprétation graphique des classifications en morphotypes, nous avons constitué un nouvel espace quantitatif de représentation de dimension 44 résumé dans le tableau 5.2. Dans la suite de ce chapitre, le traitement de ces vecteurs de caractéristiques est présenté comme un problème d'analyse factorielle servant à la réduction de dimension pour dégager des caractéristiques synthétiques et finalement pour classer les aires urbaines selon leurs spécificités d'organisation les plus structurantes.

TABLE 5.2 – Ensemble des variables composant l'espace de caractéristique des aires urbaines

Repartition de la population		Fragmentation des régions		Relations d'adjacence		Morphologie des agglomérations	
P1	Centre principal	F1	Centre principal	C11/21	Centre prin. / Cour. dense	M1	Eccentricité
P2	Centre secondaire	F2	Centre secondaire	C11/22	Centre prin. / Cour. peu dense	M2	1/4 de couronne
P3	Centre tertiaire	F3	Centre tertiaire	C11/41	Centre prin. / Grd ensemble	M3	Demi-couronne
P4	Couronne dense	F4	Couronne dense	C12/21	Centre sec. / Cour. dense	M4	3/4 de couronne
P5	Couronne peu dense	F5	Couronne peu dense	C12/22	Centre sec. / Cour. peu dense		
P6	Noyau connecté	F6	Noyau connecté	C13/22	Centre ter. / Cour. peu dense		
P7	Noyau relais	F7	Noyau relais	C13/23	Centre ter. / Noy. connecté		
P8	Noyau satellite	F8	Noyau satellite	C21/22	Cour. dense / Cour. peu dense		
P9	Noyau isolé	F9	Noyau isolé	C21/41	Cour. dense / Grd ensemble		
P10	Grands ensembles	F10	Grands ensembles	C22/41	Cour. Peu dense / Grd ensemble		
P11	Habitat dispersé	F11	ZA intégrée	C23/22	Noy. connecté / Cour. peu dense		
		F12	ZA périphérique	C23/23	Noy. connecté / Noy. Connecté		
				C23/41	Noy. Connecté / Grd ensemble		
				P12/21	Centre sec. / Cour. dense		
				P12/22	Centre sec. / Cour. peu dense		
				P13/22	Centre ter. / Cour. peu dense		
				P13/23	Centre ter. / Noy. connecté		
				P21/22	Cour. dense / Cour. peu dense		

5.2 Méthodes d'analyse factorielle

Lorsque le nombre de variables décrivant un jeu de données est important, il peut être intéressant d'avoir recours à des méthodes de réduction de dimension pour faciliter la réalisation d'une tâche d'apprentissage statistique. Ces méthodes basées sur la sélection ou l'extraction de variables trouvent leur intérêt pour des questions de temps de calcul, de visualisation des données, ou encore pour fournir une représentation plus discriminante des données. De plus, elles permettent de répondre au problème qui peut souvent être rencontré sur des jeux de données à grande dimension, qualifié de « fléau de la dimension » (*Curse of dimensionality*) par [Bellman \(1961\)](#). Étant donné que le nombre d'observations nécessaires pour construire un modèle statistiquement significatif croît de façon exponentielle avec la taille de l'espace des observations, il peut être préférable, comme dans le cas présent, de se ramener à un espace réduit basé sur des variables synthétisant les principales caractéristiques de l'espace.

Dans cette section nous nous focaliserons sur les méthodes d'extraction de variables pour réduire la taille de l'espace de caractéristiques décrivant les structures urbaines, créé dans la section précédente. Pour un jeu de données à p dimensions le problème consiste donc à identifier un sous-espace de taille $m \lll p$ à l'aide d'une fonction $y = f(x)$ permettant de transformer un ensemble de données $x = \{x_1, x_2, \dots, x_p\}$ en un nouvel ensemble de données $y = \{y_1, y_2, \dots, y_m\}$.

Plusieurs stratégies sont envisageables pour construire cette transformation. Elles peuvent être résumées en deux grandes familles : les transformations linéaires et les transformations non linéaires. Dans la famille des transformations linéaires, on retrouve

les méthodes classiques telles que l'analyse en composante principale (ACP) ou l'analyse discriminante linéaire (LDA). Celle des transformations non linéaires contient de nombreuses méthodes employant des formalismes variés, parmi lesquelles on peut citer les méthodes de *Multidimensional Scaling* (MDS), les méthodes basées sur des décompositions spectrales ou encore celles utilisant le formalisme des réseaux de neurones. Une littérature très abondante peut être trouvée sur ces deux types d'approches dans des ouvrages généralistes de data mining et d'analyses multivariées de données comme ceux de Han et Kamber (2006); Eriksson et al. (2001); Rencher (2002), ou ceux plus spécialisés de Fodor (2002) ou de Lee et Verleysen (2007) se focalisant sur les méthodes de réduction de dimension non linéaires.

Dans ce qui suit, nous nous attacherons à présenter plus en détail certaines méthodes linéaires de réduction issues de la famille des analyses factorielles. Nous choisissons ce formalisme pour obtenir un espace réduit conservant au mieux les variations présentes dans les données en se basant sur des critères de maximisation de l'inertie du jeu de données. Nous nous focaliserons ici sur les méthodes d'analyse factorielle multiple, adaptées au traitement de jeux de données structurés par groupes de variables. Nous introduirons des méthodes d'analyse factorielle multiple basées sur l'ACP et l'AFC permettant de traiter des groupes de variables quantitatives et qualitatives. Puis nous présenterons des méthodes d'analyse factorielle de données mixtes que nous utiliserons pour traiter le jeu de données des aires urbaines.

5.2.1 Les analyses factorielles multiples

Les méthodes d'analyse factorielle présentées ici sont adaptées au traitement de jeux de données structurés par groupe de variables. Dans ce cas, la matrice X des données présente la particularité d'être décomposable en blocs de matrices contenant n individus correspondant aux lignes de la matrice. En considérant que les p colonnes de X peuvent être regroupées en G groupes de variables, on note X_g la sous-matrice de taille $n \times p_g$ associée au groupe de variables g . L'objectif de ces méthodes est alors de pouvoir analyser ces G groupes de variables simultanément en prenant en compte à la fois la structure intra et inter-groupe et en cherchant à équilibrer leur influence sur l'analyse globale.

Dans notre application à la comparaison de structures urbaines, nous disposons de $G = 4$ groupes de variables parmi lesquels trois groupes sont des tableaux de fréquences et un groupe correspond à des variables quantitatives. Dans ce contexte, les méthodes d'analyse de données et de réduction de dimension à employer doivent être capables de tenir compte de la nature des variables en plus de fournir un équilibre entre groupes. Nous présentons donc dans un premier temps les méthodes d'analyse factorielle multiple permettant de traiter des groupes de variables tous quantitatifs ou qualitatifs, qui serviront de base aux méthodes étendues pour traiter des groupes de données mixtes.

5.2.1.1 L'analyse factorielle multiple (AFM) sur données quantitatives

L'analyse factorielle multiple (AFM) présentée par [Escoffier et Pagès \(2008\)](#) est une méthode de réduction de dimension liée à l'ACP, qui permet de prendre en compte la structure par groupe des variables lorsque chaque individu est caractérisé par plusieurs groupes de variables de nature quantitative. L'objectif est alors de construire une projection des données dans un espace de plus petite dimension en équilibrant l'influence que peuvent avoir les différents groupes.

Rappelons d'abord que l'ACP est une méthode linéaire de réduction de dimension introduite par les articles de [Pearson \(1901\)](#) et [Hotelling \(1933\)](#). L'objectif est de produire une transformation où les nouvelles variables, appelées composantes principales, sont des combinaisons linéaires ordonnées des variables d'entrée qui maximisent les différences entre individus. D'un point de vue géométrique, l'ACP cherche le sous-espace vectoriel de dimension $m < p$ qui ajuste au mieux la forme du nuage de points de données au sens des moindres carrés. Le nuage projeté dans le sous-espace doit avoir une inertie (variance) maximale. Cela revient dans la pratique à diagonaliser la matrice V de variance-covariance des données, et donc à rechercher la base de vecteurs propres et valeurs propres associée, telle que :

$$Vu = \lambda u \quad (5.7)$$

En AFM comme en ACP, si les variables ont des ordres de grandeur différents elles doivent être centrées-réduites afin d'être rendues comparables. Dans le cas de l'AFM, elles doivent en plus être pondérées de façon à ce qu'un groupe de variables ne domine pas les autres lors de la projection des données. La transformation proposée par l'AFM se résume alors en deux étapes décrites ci-dessous :

1. On réalise des ACP sur les G groupes de variables, pris indépendamment, pour obtenir deux informations :
 - les composantes principales $CL_k^g(l) = X_g u_l$ de valeurs propres non nulles, qui peuvent remplacer les variables initiales dans la suite de l'analyse,
 - la valeur propre $\lambda_g(1)$ associée à la première composante principale, qui va servir à pondérer chaque variable du g -ième groupe.
2. On réalise une ACP sur l'ensemble des groupes de variables pris simultanément où les données en entrée sont transformées de la façon suivante :
 - la matrice X est conservée telle quelle, ou elle est remplacée par la matrice C des composantes principales CL_k^g de chaque groupe de variables,
 - le poids $1/\lambda_g(1)$ est affecté à chaque variable du g -ème groupe.

Associer à chaque groupe de variable un coefficient unique de valeur $1/\lambda_g(1)$ assure que les formes des sous-nuages de points sont préservées, mais aussi que les sous-nuages

de points sont normalisés : l'inertie d'un groupe sur un axe quelconque vaut au maximum 1. Dans la pratique, l'étape 2 de l'ACP pondérée peut être réalisée de façon équivalente sur l'ensemble des groupes de variables initiales ou sur les composantes principales des ACP marginales puisque ces composantes principales partielles correspondent aux projections des individus dans une base orthonormée. Les nuages de points associés aux variables initiales ou aux composantes principales sont équivalents. Cette propriété devient intéressante lorsque les groupes de variables sont de grande dimension puisqu'il sera alors possible de réaliser l'AFM uniquement sur les composantes principales associées à des valeurs propres supérieures à un certain seuil, ou encore sur les composantes principales considérées comme suffisamment représentatives du jeu de données. L'AFM qui en découlera sera une approximation de l'AFM qui aurait été réalisée sur les données brutes, dont la qualité dépendra du choix des composantes retenues.

L'interprétation des résultats d'une AFM est basée sur les moyens d'interprétations de l'ACP, c'est-à-dire sur l'étude de l'inertie de chaque axe donnée par les valeurs propres, et sur les contributions des individus et des variables calculées à partir des vecteurs propres. De plus [Escoffier et Pagès \(2008\)](#) mentionnent plusieurs aides supplémentaires à l'interprétation des résultats d'une AFM, basées sur l'étude de l'influence des variables prises en tant que groupe. Il est ainsi possible de calculer les contributions des g groupes de variables à chaque composante principale k :

$$Ctr_k[g] = \sum_{j \in X_g} Ctr_k(j) \quad (5.8)$$

où $Ctr_k(j)$ est la contribution d'une variable initiale j à une composante principale k donnée par :

$$Ctr_k(j) = \frac{u_{j,k}^2}{\lambda_g(1)} \quad (5.9)$$

et où $u_{j,k}$ est le vecteur propre associé à la k -ième composante principale.

Il est aussi possible de calculer les corrélations entre les différents groupes de variables à l'aide d'un coefficient de liaison tel que :

$$L_{(g,g')} = \frac{\text{trace}(X_g X_g^t \times X_{g'} X_{g'}^t)}{\lambda_k(1) \times \lambda_{g'}(1)} \quad (5.10)$$

Ce même coefficient peut être employé pour mesurer la corrélation entre une composante principale et un groupe de variables.

5.2.1.2 Extensions de l'AFM aux données qualitatives ou aux tableaux de fréquences

L'application de l'AFM à des données qualitatives est conditionnée par la façon dont peuvent être représentés les groupes de données. Si les données peuvent être codées sous forme disjonctive complète, alors la méthode d'Analyse des Correspondances

Multiples (ACM) peut être utilisée pour traiter les données. [Escoffier et Pagès \(2008\)](#) ont montré que la réalisation d'une ACM est équivalente à l'application d'une ACP avec les pondérations adéquates sur les indicatrices, ce qui permet de définir une extension de l'AFM dans le cas où les groupes de données sont qualitatifs et avec un codage binaire. L'extension de l'AFM aux tableaux de contingence ou de fréquences, traités dans le cas non groupé par analyse factorielle des correspondances (AFC), est plus délicate à mettre en œuvre.

Rappelons ici que l'AFC a été développée sous ce nom par [Benzécri \(1973\)](#) pour étudier des variables organisées sous forme de tableaux de contingence contenant les effectifs ou fréquences croisés des modalités de deux variables qualitatives. Le but de l'AFC est en fait de trouver la transformation qui fournit la meilleure représentation simultanée des deux variables qualitatives considérées. On considère alors un tableau de fréquence F de terme général f_{ij} :

$$f_{ij} = \frac{x_{ij}}{n} \quad \text{où} \quad n = \sum_{i,j} x_{ij} \quad (5.11)$$

A cette matrice de fréquences relatives F sont associés deux vecteurs $f_{i.}$ et $f_{.j}$ de fréquences marginales pour les lignes (les individus) et les colonnes (les variables) tels que :

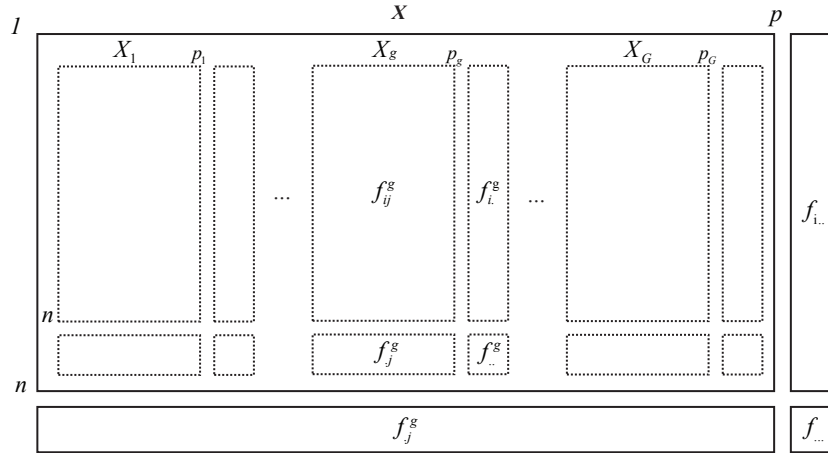
$$f_{i.} = \sum_j f_{ij} \quad \text{et} \quad f_{.j} = \sum_i f_{ij} \quad (5.12)$$

L'AFC est une variante de l'ACP qui diffère simplement par la nature des données considérées en entrée et donc par l'utilisation d'une métrique différente de la métrique euclidienne pour mesurer la proximité entre individus : la métrique du χ^2 . Dans la pratique les composantes principales sont trouvées de la même façon qu'en ACP, c'est-à-dire en diagonalisant de manière équivalente la matrice d'inertie du nuage de points du profil-ligne $f_{ij}/f_{i.}$ ou du profil-colonne $f_{ij}/f_{.j}$, pondérée par les fréquences marginales $f_{i.}$ ou $f_{.j}$, et à laquelle on associe la métrique $f_{.j}$ ou $f_{i.}$.

L'application simultanée de l'AFC à plusieurs tableaux est donc compliquée par la nécessité d'affecter à chaque ligne un poids variable correspondant dans le cas simple à la fréquence marginale par ligne $f_{i.}$. Le traitement simultané de plusieurs groupes ne peut alors être fait que s'il est possible de définir de nouvelles fréquences marginales qui fassent toujours sens pour mesurer la ressemblance des individus par la distance du χ^2 .

Plusieurs méthodes ont été développées pour traiter ce problème, mais nous nous intéressons ici à celles qui gardent la métrique de χ^2 de l'AFC et qui intègrent aussi le formalisme de l'AFM, c'est-à-dire la prise en compte à la fois des différences intra et de l'équilibre inter groupes. Deux méthodes retiennent alors notre attention : l'Analyse Simultanée (AS) de [Zàrraga et Goitisolo \(2002\)](#) et l'Analyse Factorielle Multiple de

FIGURE 5.6 – Ensemble des matrices traitées par AFM et notations



tableaux de contingence (AFMTC) de Bécue-Bertaut et Pagès (2004). Ces deux méthodes très proches, qui sont comparées dans (Zàrraga et Goitisoló, 2009), diffèrent par la façon de concaténer les tableaux et donc par la pondération choisie pour remplacer les fréquences marginales f_i , d'une AFC simple. L'AS considère les matrices partielles F_1, F_2, \dots, F_G et leurs marges partielles f_i^g , tandis que l'AFMTC considère la matrice globale F et ses marges globales $f_{i\cdot}$ (cf. figure 5.6) pour réaliser l'analyse. Elles reprennent la structure générale de l'AFM en deux étapes : les analyses partielles de chaque groupe sont réalisées pour obtenir les pondérations adéquates ; et l'analyse globale de tous les groupes est effectuée en tenant compte de ces pondérations. Pour les deux méthodes et les deux étapes, les analyses effectuées correspondent à des AFC appliquées en utilisant le formalisme de l'ACP. Le tableau 5.3 résume les méthodologies et différences entre ces deux méthodes.

Ces variantes possèdent des propriétés justifiant leur intérêt respectif. Lorsque les fréquences marginales des tableaux de contingence sont similaires pour tous les groupes, alors les deux méthodes fournissent des résultats équivalents puisque les fréquences marginales partielles et globales sont similaires. Les différences apparaissent donc lorsque les fréquences marginales partielles ne sont plus proportionnelles à la fréquence marginale globale. Parmi les propriétés remarquables, on peut retenir que :

1. l'AS prend en compte les poids et métriques des analyses partielles des G groupes : l'analyse globale est donc plus proche des analyses partielles que l'analyse globale obtenue par AFMTC,
2. l'AS peut être appliquée sur des tableaux de contingence qui correspondent à des groupes de variables qui n'ont pas nécessairement des fréquences marginales identiques,
3. l'AFMTC fournit une représentation centrée des profils-lignes, ce qui facilite l'analyse sachant que le barycentre correspond au profil moyen autour duquel se dispersent les individus.

Les éléments d'interprétation de ces analyses multiples sont les mêmes que pour une AFC classique, à savoir l'inertie des axes, ou les contributions ou les qualités de représentation des lignes ou des colonnes. Nous ne donnons ici que les formules de calcul des aides à l'interprétation de l'AS, mais les aides à l'interprétation de l'AFMTC peuvent être déduites de ces formules en reprenant les pondérations et matrices données dans le tableau 5.3. Les contributions des lignes et des colonnes aux axes de AS sont données par :

$$Ctr_{ik} = \frac{p_{ic} CL_k(i^c)^2}{\lambda_k} \quad \text{et} \quad Ctr_{jk} = \frac{f_{.j}^g CC_k(j)^2}{\lambda_k} \quad (5.13)$$

Les qualités de représentations sont données par :

$$\cos(i_g, k) = \frac{CL_k(i^c)^2}{d^2(i^c, 0)} \quad \text{et} \quad \cos(j, k) = \frac{CC_k(j)^2}{d^2(j, 0)} = \frac{CC_k(j)^2}{\sum_{k'} CC_{k'}(j)^2} \quad (5.14)$$

En plus de ces mesures issues de l'AFC classique, il est possible de mesurer les contributions par groupe comme en AFM (équation (5.8)) ou de mesurer les corrélations entre groupes de variables ou entre composantes principales et groupes de variables en utilisant l'équation (5.10). Les représentations simultanées des composantes partielles et globales peuvent aussi servir d'aide à l'interprétation de l'influence de chaque groupe sur la projection globale des données.

5.2.2 Extensions de l'AFM aux données mixtes

Le traitement de données mixtes (mesures quantitatives et qualitatives, ou tableaux de fréquence) est un problème largement documenté qui regroupe diverses approches et applications (Creswell et Clark, 2010). Dans le cas de l'extension de l'AFM au traitement de données mixtes, la problématique est de devoir définir une distance globale qui puisse rendre compte de la proximité entre individus aux caractéristiques de nature quantitative ou qualitative. Nous nous intéressons ici aux travaux de Bécue-Bertaut et Pagès (2008) qui ont étendu le formalisme de l'AFM au traitement de données mixtes. Des données de différentes natures peuvent toutes être traitées par des méthodes d'analyses factorielles : ACP pour les données quantitatives, ACM pour les données qualitatives au codage binaire et AFC pour les données qualitatives sous forme de tableau de contingence. Ces trois variantes de l'analyse factorielle mesurent la proximité entre les données selon des distances différentes : la distance euclidienne pour l'ACP, et la distance du χ^2 pour l'AFC et l'ACM. Leur extension de l'AFM repose donc sur le traitement des données mixtes en utilisant une distance globale résultant de la combinaison des distances partielles induites par la nature des données.

TABLE 5.3 – Méthodologies comparées de l'AS et l'AFMTC

Analyse Simultanée (AS)		AFM sur Tableaux de Contingence (AFMTC)		
1 ^{ÈRE} ÉTAPE	Analyse des matrices partielles			
	Profil-ligne $i^g = \left(\frac{f_{ij}^g}{f_{i.}^g} - f_{.j}^g \right)$	Poids $f_{i.}^g$	Profil-ligne $i^g = \left(\frac{f_{ij}^g}{f_{i..}} - f_{.j}^g \right)$	Poids $f_{i..}$
	Profil-colonne $j^g = \left(\frac{f_{ij}^g}{f_{.j}^g} - f_{i.}^g \right)$	Poids $f_{.j}^g$	Profil-colonne $j^g = \left(\frac{f_{ij}^g}{f_{.j}^g} - \frac{f_{i.}^g}{f_{..}^g} \right)$	Poids $f_{.j}^g$
	Matrice à diagonaliser			
	Profil-ligne $T = Y^t Y$ $y_{ij}^g = \frac{f_{ij}^g - f_{i.}^g f_{.j}^g}{\sqrt{f_{i.}^g f_{.j}^g}}$	Profil-colonne $U = Y Y^t$	Profil-ligne $T = Y^t Y$ $y_{ij}^g = \frac{f_{ij}^g - \frac{f_{i.}^g f_{.j}^g}{f_{..}^g}}{\sqrt{f_{i..} f_{.j}^g}}$	Profil-colonne $U = Y Y^t$
Sorties :				
Valeurs propres $(\lambda_1(1), \dots, \lambda_g(1), \dots, \lambda_G(1))$ de la décomposition spectrale de T ou U				
Pondérations : $\alpha_g = \frac{1}{\lambda_g(1)}$ ou $\alpha_g = \frac{\sum_g \lambda_g(1)}{\lambda_g(1)}$				
2 ^{ÈME} ÉTAPE	Analyse de la matrice globale			
	Profil-ligne $i^c = \sum_{g \in \mathcal{G}} \frac{\sqrt{f_{i.}^g}}{\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g}} i^g$	Poids $p_{i^c} = \left(\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g} \right)^2$	Profil-ligne $i^c = \left(\frac{f_{ij}^g}{f_{i..}} - f_{.j}^g \right)$	Poids $f_{i..}$
	Profil-colonne $j = \frac{\sqrt{\alpha_g}}{\sqrt{f_{.j}^g}} \left(\frac{f_{ij}^g}{f_{.j}^g} - f_{i.}^g \right)$	Poids $f_{.j}^g$	Profil-colonne $j = \left(\frac{f_{ij}^g}{f_{.j}^g} - \frac{f_{i.}^g}{f_{..}^g} \right)$	Poids $\alpha_g f_{.j}^g$
	Matrice à diagonaliser			
	Profil-ligne $V = Z^t Z$ $z_{ij} = \sqrt{\alpha_g} \frac{f_{ij}^g - f_{i.}^g f_{.j}^g}{\sqrt{f_{i.}^g f_{.j}^g}}$	Profil-colonne $W = Z Z^t$	Profil-ligne $V = Z^t Z$ $z_{ij} = \sqrt{\alpha_g} \frac{f_{ij}^g - \frac{f_{i.}^g f_{.j}^g}{f_{..}^g}}{\sqrt{f_{i..} f_{.j}^g}}$	Profil-colonne $W = Z Z^t$
Sorties :				
Décompositions spectrales : $V v_k = \lambda_k v_k$ et $W w_k = \lambda_k w_k$				
Composantes principales globales				
$CL_k(i^c) = \frac{1}{\sqrt{p_{i^c}}} \sum_j z_{ij} v_{jk}$		$CL_k(i^c) = \frac{1}{\sqrt{f_{i..}}} \sum_j z_{ij} v_{jk}$		
$CC_k(j) = \frac{1}{\sqrt{f_{.j}^g}} \sum_i z_{ij} w_{ik} = \sqrt{\frac{\lambda_k}{f_{.j}^g}} v_{jk}$		$CC_k(j) = \frac{1}{\sqrt{f_{.j}^g}} \sum_i z_{ij} w_{ik} = \sqrt{\frac{\lambda_k}{f_{.j}^g}} v_{jk}$		
Composantes principales partielles				
$CL_k(i^g) = \frac{\sqrt{\alpha_g}}{\sqrt{f_{i.}^g}} \sum_j y_{ij}^g v_{jk}$		$CL_k(i^g) = \frac{1}{\sqrt{\lambda_k}} \frac{f_{i.}^g}{f_{i..}} \sum_j \frac{f_{ij}^g}{f_{i.}^g} \alpha_g CC_k(j, g)$		

N.B. : les poids dans l'espace des profils-ligne correspondent aux métriques dans l'espace des profils-colonne et vice-versa (*Saporta, 1990*).

La réalisation d'une AFM sur des données mixtes consiste en l'application d'une ACP pondérée (centrée mais non réduite) sur la matrice contenant les variables de différentes natures. Apparaît alors le problème de choisir les pondérations adéquates. Les colonnes sont pondérées de façon similaire à l'AFM, c'est-à-dire par $\frac{1}{\lambda_g(1)}$ l'inverse de la première valeur propre de chaque groupe multipliée par la pondération initiale des colonnes utilisées dans les analyses partielles. La pondération des lignes est moins triviale compte tenu des poids différents qui seraient utilisés dans les analyses partielles. La pondération est uniforme, du type $p_i = \frac{1}{n}$ lorsque les analyses partielles concernent des variables quantitatives et binaires (ACP et ACM). Mais elle est non uniforme, du type fréquence marginale $f_{i.}$ ou $f_{.i.}$ lorsque les analyses partielles concernent un ou plusieurs tableaux de fréquence. La solution proposée est alors de choisir une des deux pondérations, uniforme ou non uniforme, et de l'appliquer à l'ensemble des groupes de variables du jeu de données quelle que soit leur nature. En conséquence la distance entre individus pour les variables fréquentielles ou pour les variables quantitatives et binaires sera modifiée, ce qui modifiera l'inertie maximisée. Le choix de la pondération sur les lignes doit donc être fait en fonction des objectifs de l'analyse.

L'application que nous souhaitons réaliser concerne un jeu de données composé de variables majoritairement fréquentielles divisées en trois groupes concernant la répartition de la population, du nombre de régions et de paires de régions de morphotypes donnés, et d'un groupe de variables quantitatives concernant la forme des agglomérations principales. La nature de ces variables appelle à quelques adaptations de la méthode proposée par [Bécue-Bertaut et Pagès \(2008\)](#). En effet pour déterminer les principales tendances d'organisation spatiale des aires urbaines, nous proposons d'utiliser une méthode d'analyse factorielle multiple de données mixtes (AFMM) mais avec les deux adaptations suivantes :

- Variables utilisées dans l'analyse globale : CL_k^g , les composantes principales de chaque analyse partielle.
- Pondération des lignes : $p_i = p_{i.}$, la pondération des profils-lignes de l'AS.

Le choix de la pondération de l'AS résulte de la nature des données que nous souhaitons traiter. La majorité des variables peuvent être exprimées comme des mesures fréquentielles de population, de nombre de régions ou de paires de régions par catégorie de morphotype. Seules les variables de forme des agglomérations principales sont des variables quantitatives ne représentant pas des fréquences. Ainsi il semble important de conserver la pondération des AFC pour le calcul de la distance globale entre individus. La pondération de l'AS est préférée à celle de l'AFMTC puisque les trois groupes de variables fréquentielles correspondent à des comptages d'objets de natures différentes. Ainsi rien n'assure que les fréquences marginales pour les individus de chaque groupe seront identiques ou même proportionnelles : l'AS est plus adaptée que l'AFMTC.

Le choix de travailler sur les composantes principales partielles plutôt que sur les

variables d'origine résulte lui de la dimension importante de chaque groupe de variables. Appliquer l'analyse globale sur les composantes principales les plus représentatives de chaque groupe permettrait de réduire préalablement la dimensionnalité des données, pour ne conserver que les variables synthétisant les plus importantes différences entre les aires urbaines. Nous montrons qu'en reprenant les pondérations de l'AS, il est aussi possible d'appliquer l'AFMM de façon équivalente sur la matrice X des données initiales ou sur les composantes principales partielles. En AS la pondération du profil-ligne est choisie de façon à ce que l'inertie globale du jeu de données s'exprime comme la somme pondérée des inerties partielles :

$$Inertie = \sum_{g \in \mathcal{G}} \alpha_g Inertie[i^g] \quad (5.15)$$

Sachant que l'inertie d'un jeu de données peut être exprimée selon les composantes principales d'une ACP simple ou d'une AFC simple de la manière suivante :

$$Inertie = \sum_k \lambda_k = \begin{cases} \frac{1}{n} \sum_{k,i} CL_k(i)^2 & \text{(ACP)} \\ \sum_{k,i} f_{i.} CL_k(i)^2 & \text{(AFC)}, \end{cases} \quad (5.16)$$

on peut décomposer l'inertie globale du jeu de données mixtes en fonction des groupes de variables fréquentielles $g \in \mathcal{G}_f$ et des groupes de variables quantitatives $g \in \mathcal{G}_q$:

$$Inertie = \sum_{g \in \mathcal{G}_f} \alpha_g \sum_{i,k} f_{i.}^g CL_k^g(i)^2 + \sum_{g \in \mathcal{G}_q} \frac{\alpha_g}{n} \sum_{i,k} CL_k^g(i)^2 \quad (5.17)$$

Dans ce cas, la matrice à diagonaliser $V = L^t L$ pour obtenir la projection du profil-ligne global est identifiée à partir de la matrice L de terme général :

$$l_{ij} = \begin{cases} \sqrt{\alpha_g} \sqrt{f_{i.}^g} CL_j^g(i) & \text{si } j \in \mathcal{G}_f \\ \frac{\sqrt{\alpha_g}}{\sqrt{n}} CL_j^g(i) & \text{si } j \in \mathcal{G}_q \end{cases} \quad (5.18)$$

Les composantes principales globales pour les lignes sont identiques à celles obtenues par l'analyse des données originales. Elles sont données en fonction des vecteurs propres v_k de la matrice V :

$$CL_k(i^c) = \frac{1}{\sqrt{p_{i^c}}} \sum_j l_{ij} v_k \quad (5.19)$$

Pour les colonnes, les composantes principales issues de cette analyse réalisée sur les facteurs partiels ne sont pas les mêmes que celles qui seraient obtenues si l'analyse était réalisée sur les données originales. Le nombre de composantes principales issues des AFC partielles vaut au maximum $p_g - 1$ (Saporta, 1990), ainsi il n'y a pas le même nombre de variables initiales que de composantes principales partielles, et les composantes principales globales ne sont pas exprimées dans le même espace projeté. La dualité des analyses des lignes et des colonnes étant une propriété propre aux analyses factorielles (Escoffier et Pagès, 2008, Chapitre 5), les composantes principales des colonnes peuvent être retrouvées avec la formule suivante :

$$CC_k(j) = \frac{1}{\sqrt{\lambda_k}} \sum_i \frac{1}{\sqrt{p_{i^c}}} l_{ij} CL_k(i^c) \quad (5.20)$$

Les aides à l'interprétation de cette analyse factorielle multiple sont alors les mêmes que celles d'une ACP ou d'une AFC pour les individus et les variables prises indépendamment et d'une AFM pour les groupes de variables.

La méthode d'AFMM décrite ici nous permettra donc d'analyser les différentes caractéristiques morphologiques marquant l'organisation des aires urbaines. L'avantage de cette approche est de pouvoir regrouper dans une analyse globale les résultats des analyses partielles réalisées par AFC ou ACP. Nous verrons dans la suite que se baser sur ces analyses partielles nous permettra de proposer une première phase de réduction de dimension qui constituera une aide pour l'interprétation des différentes tendances d'organisations.

5.3 Identification des tendances spatialement structurantes

Par la construction de vecteurs de caractéristiques et par l'emploi de méthodes d'analyse factorielle, nous cherchons à définir les tendances différenciant les structures urbaines construites à partir d'informations sur les densités de population et de surface bâtie. Pour ce dernier travail à l'échelle de la macro-forme (celle de l'image), nous explorons les tendances d'organisation urbaine à partir de caractéristiques résumant des groupes de variables composés d'indices de formes des agglomérations, de tableaux de répartition de population par morphotype, du nombre de régions par morphotype et du nombre de paires de régions adjacentes de morphotypes distincts.

Pour identifier ces caractéristiques marquantes nous proposons d'utiliser la méthode d'analyse factorielle multiple de données mixtes (AFMM) présentée dans la section précédente. Nous pourrions alors interroger les traits communs des structures urbaines étudiées et chercher à définir des grands types d'organisations en se basant sur l'étude des proximités entre structures représentées par ces caractéristiques à l'aide d'une CAH. Ici nous travaillons à partir de méthodes d'analyse statistique toutes non spatiales puisque l'espace de représentation des aires urbaines intègre cette composante spatiale dans les vecteurs de caractéristiques pris en entrée. Les catégories de morphotypes et agglomérations sur lesquels se basent toutes les mesures représentent des entités spatialement déterminées. Elles sont associées à des positions plus ou moins centrales ou périphériques au sein de l'aire urbaine qui se retrouvent intégrées dans ces différentes modalités qualitatives.

Nous présentons donc dans un premier temps les résultats des analyses partielles sur lesquelles nous nous basons pour ne sélectionner qu'un nombre restreint de composantes principales résumant les caractéristiques explicatives de l'organisation et des oppositions entre aires urbaines. Ces composantes principales sont ensuite utilisées

pour représenter chaque groupe de variable dans l'analyse globale assurant un traitement équilibré de l'information apportée par chaque groupe de variables. Enfin nous présentons les résultats de l'application d'une CAH permettant d'obtenir les différents types d'organisation observables parmi les 50 aires urbaines de l'échantillon d'étude.

5.3.1 Les analyses partielles des groupes de variables

En suivant une démarche analytique, nous faisons le choix de présenter dans un premier temps les analyses factorielles partielles de chaque groupe de variables. L'idée est ici de chercher à exprimer des espaces partiels de description des aires urbaines en termes de composantes principales fournissant un nombre réduit de facteurs conservant un grand pouvoir explicatif. Cette phase intermédiaire de traitement nous permet d'établir un premier niveau d'analyse et d'interprétation des différentes caractéristiques d'organisation, qui facilitera l'analyse des caractéristiques globales constituées à partir d'un nombre plus réduit de variables.

5.3.1.1 Analyse de la morphologie des agglomérations principales

Le jeu de données étudié dans cette première analyse agrège les quatre variables quantitatives mesurant les caractéristiques de forme des agglomérations principales des 50 aires urbaines : l'excentricité et les ratios de surface répartie dans un quadrant centré sur ces agglomérations.

TABLE 5.4 – Décomposition de l'inertie par axe de l'ACP sur indices de forme

	C1	C2	C3	C4
Valeurs propres	2,99	0,78	0,11	0,05
% d'inertie	76%	20%	3%	1%
% d'inertie cumulée	76%	96%	99%	100%

Le tableau 5.4 permet d'étudier la forme générale du jeu de données à travers l'étude de la répartition de l'inertie sur les axes de l'ACP normée. Il révèle que le premier axe de l'ACP est associé à une forte valeur propre ($\lambda_1 = 3.01$) ce qui traduit la forte corrélation entre les variables du jeu de données. Cette corrélation est liée à la construction volontairement dépendante des indices de forme de développement des couronnes. Ce facteur très synthétique représente à lui seul 75% de l'inertie totale du jeu de données. Compte tenu de la très faible part d'inertie représentée par les deux derniers facteurs, ils sont écartés du reste de l'analyse.

La figure 5.7 représentant le cercle des corrélations dans le premier plan factoriel, ainsi que la projection des aires urbaines dans ce même plan, permet de donner une signification aux composantes principales. Elle révèle que les quatre variables initiales

contribuent de façon équilibrée à la création du premier axe factoriel. Les corrélations toutes du même signe et du même ordre de grandeur indiquent que le premier axe définit un facteur de « taille » : le facteur est approximativement proportionnel à la moyenne des variables initiales (Saporta, 1990). Le second axe est lui fortement lié à l'excentricité qui contribue très majoritairement à sa création. Il constitue en complément du premier un facteur de « forme » qui classe les aires urbaines de même moyenne en fonction de leur excentricité.

En analysant les projections des individus dans le premier plan factoriel il est possible d'identifier les aires urbaines représentatives des différentes tendances de forme. Notamment pour le premier axe, les aires urbaines aux coordonnées les plus négatives sont celles qui possèdent en moyenne des valeurs élevées pour les quatre variables initiales, traduisant ainsi une forme linéaire et surtout un développement contraint dans 1/4, 1/2 ou 3/4 de couronne. Le développement contraint en 1/4 de couronne de l'agglomération de Bayonne est une caractéristique qui la singularise et lui confère un poids important sur l'axe. La forte excentricité, marqueur de forme linéaire, des agglomérations de Nice, Saint-Etienne, Saint-Nazaire et Marseille constitue une deuxième source de différenciation. A l'opposé, les aires urbaines aux coordonnées positives et aux positions les plus extrêmes (Tours, Poitiers, Lyon) peuvent être interprétées comme des formes au développement non contraint dans une direction particulière.

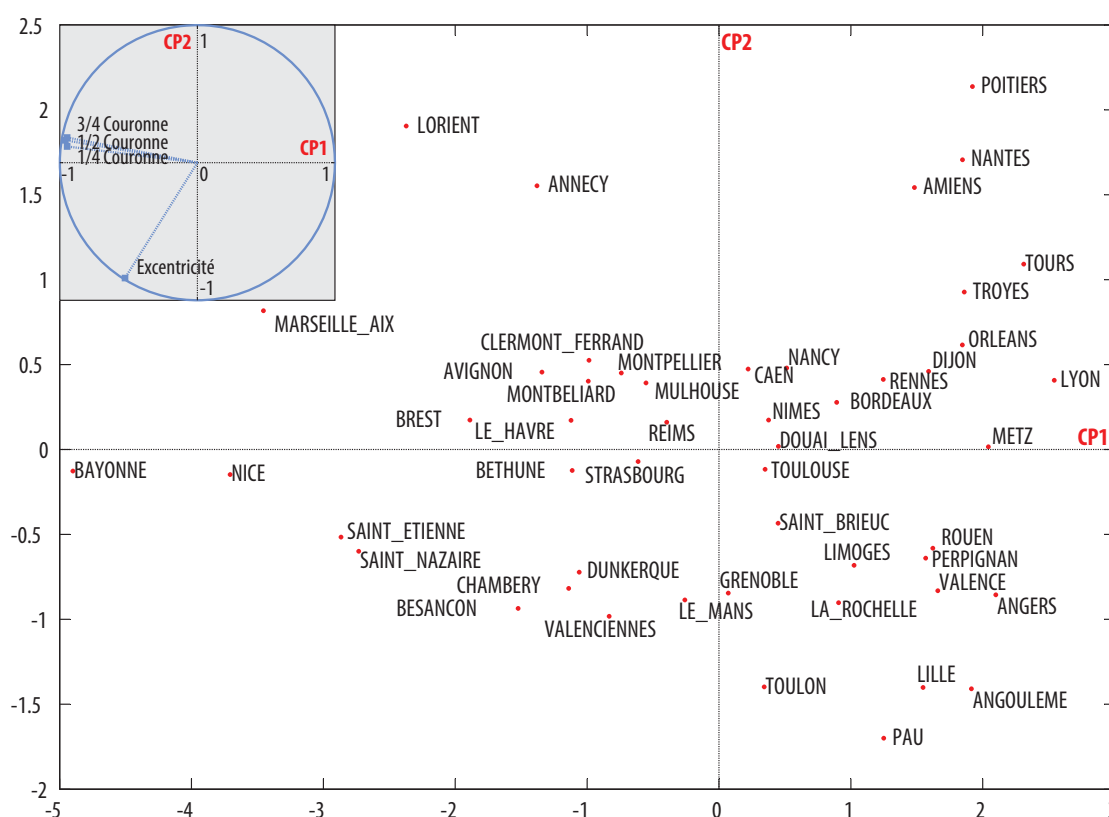


FIGURE 5.7 – Premier plan factoriel de l'ACP sur les indices de forme

Même s'il apporte peu en termes d'inertie, le deuxième axe est intéressant pour compléter la différenciation des formes de développement des agglomérations selon leur excentricité. Parmi les aires urbaines dont la coordonnée est fortement positive, on distingue nettement le groupe de celles qui possèdent un développement isotrope circulaire (Poitiers, Nantes, Amiens), de celles qui possèdent une forme globalement circulaire mais un développement contraint par des obstacles dans une ou plusieurs directions de développement (Lorient, Annecy). L'opposition est moins marquée pour les agglomérations avec des coordonnées négatives comme Pau, Angoulême, Toulon ou Lille. Toutes sont caractérisées par une forme elliptique, associée à un développement qui s'est fait de façon relativement symétrique.

5.3.1.2 Analyse de la répartition de la population

Le jeu de données étudié ici concerne la répartition de la population au sein des aires urbaines. Les 11 variables en entrée correspondent au dénombrement de la population pour chaque catégorie de morphotype. Même si les données analysées ne forment pas véritablement un tableau à double entrée, nous nous plaçons ici dans le formalisme de l'AFC pour analyser la répartition de la population. Les catégories de morphotype qui représentent les modalités d'une variable de forme de développement, et les aires urbaines qui représentent les individus (ou les modalités d'une variable ville) forment en effet un tableau de données qui se transforme naturellement en un tableau de fréquences.

TABLE 5.5 – Décomposition de l'inertie par axe de l'AFC répartition de population

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Valeurs propres	0,08	0,07	0,03	0,02	0,02	0,01	0,01	0,01	0,01	0,00
% d'inertie	29%	26%	13%	9%	8%	5%	4%	3%	2%	1%
% d'inertie cumulée	29%	55%	68%	77%	84%	90%	94%	97%	99%	100%

L'analyse des résultats de l'AFC à partir de la répartition de l'inertie sur les différents axes (tableau 5.5) montre qu'il y a plusieurs directions d'inertie d'importance à peu près équivalente. Il ne semble pas y avoir de corrélations fortes entre les variables et les individus puisque toutes les valeurs propres (qui varient entre 0 et 1) sont assez faibles. Il faut considérer 5 composantes sur 10 pour conserver au moins 80% de l'inertie du jeu de données. Nous conserverons pour la suite de l'analyse les cinq premières composantes principales, les autres pouvant être considérées comme négligeables.

La figure 5.8 présente la projection simultanée des aires urbaines et des variables initiales dans le premier plan factoriel de cette AFC. Le premier axe ordonne les aires urbaines en fonction de la part plus ou moins importante de population en centre principal ou secondaire. Avec de fortes contributions et coordonnées positives, les aires

urbaines de Lille, Nice, Marseille-Aix constituent les représentantes les plus marquantes de cette tendance. Les aires urbaines possédant une part importante de population dans des noyaux isolés ou de l'habitat dispersé sont opposées à cette tendance sur le premier axe. Avec leurs fortes contributions et coordonnées négatives, Poitiers, Besançon ou Saint-Brieuc en sont les représentantes. L'axe 2 oppose d'une part des aires urbaines à la population importante dans les noyaux connectés et couronnes peu denses (Douai-Lens, Mulhouse, Valenciennes et Béziers), et d'autre part des aires urbaines contenant une part importante de population en noyaux isolés.

L'analyse de la constitution des autres composantes principales peut être effectuée de la même façon en se basant sur les projections des individus et des variables, mais aussi sur leurs contributions. Sans être nécessairement mal représentées, toutes les modalités de morphotypes ne contribuent pas de façon significative à la création des premiers axes. Certaines modalités relativement rares et variant peu entre les différentes aires urbaines, comme la population en grands ensembles ou en noyaux relais, ont un faible pouvoir explicatif dans la différenciation de l'échantillon.

Du point de vue des individus, on constate que plusieurs aires urbaines contribuent de façon significative à la création de ces 5 axes. Le second axe est un peu particulier

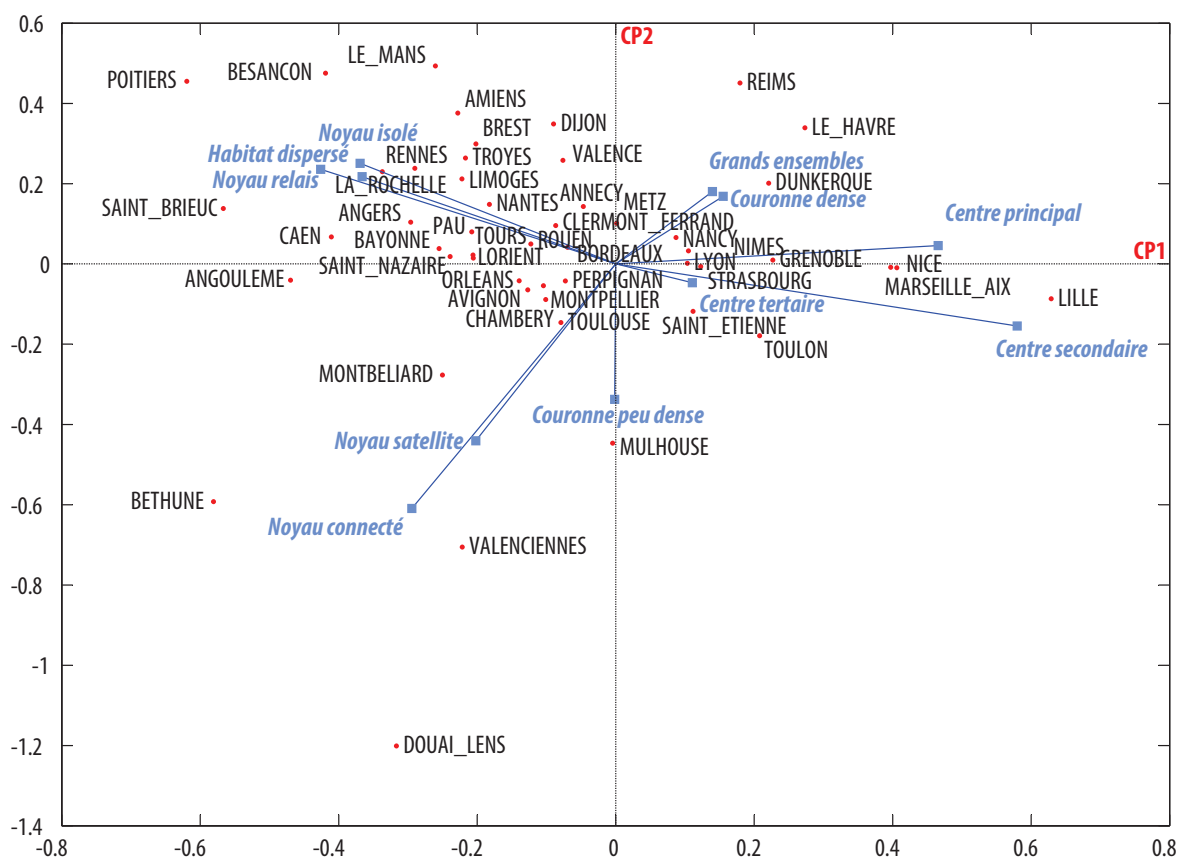


FIGURE 5.8 – Premier plan factoriel de l'AFC sur la répartition de la population

puisque l'aire urbaine de Douai-Lens contribue très fortement à la création de l'axe, ce qui se traduit par une coordonnée fortement négative. Sa singularité d'organisation marquée par une très forte population en couronnes peu denses et en noyaux connectés génère une variance élevée et un profil distant des autres aires urbaines même si on retrouve d'autres aires urbaines avec les mêmes caractéristiques d'organisation, mais dans une moindre mesure.

On retiendra de cette AFC que les cinq premières composantes peuvent résumer les tendances de répartition de la population parmi les onze morphotypes résidentiels. Les caractéristiques marquantes identifiées ici sont :

- C_{pop}^1 : Une opposition entre des ***structures urbaines concentrées*** à la population importante en centres principaux et secondaires (Lille, Aix-Marseille, Nice), et des ***structures urbaines déconcentrées*** à la population importante en noyaux isolés et en habitat dispersé (Poitiers, Saint-Brieuc).
- C_{pop}^2 : Une opposition entre des structures urbaines aux ***agglomérations peu denses*** où la population est importante dans les noyaux connectés et dans les couronnes peu denses (Douai-Lens, Valenciennes), et des structures urbaines aux ***agglomérations denses et à la périphérie développée*** où la population est importante dans les couronnes denses et les noyaux isolés (Le Mans, Reims, Besançon).
- C_{pop}^3 : Une opposition entre des structures urbaines aux ***centres très denses*** où la population est importante dans le centre principal (Lille, Strasbourg), et des structures urbaines aux ***centres moyennement denses*** où la population est importante dans les centres tertiaires et couronnes denses (Toulon, Marseille).
- C_{pop}^4 : Une opposition entre des structures urbaines au ***développement périurbain dispersé*** où la population est importante en habitat dispersé (Pau, Angoulême), et des structures urbaines aux ***agglomérations denses et à la périphérie développée*** où la population est importante dans les couronnes denses et les noyaux isolés (Caen, La Rochelle).
- C_{pop}^5 : Une opposition entre des structures urbaines aux ***agglomérations peu denses*** où la population est importante dans les couronnes peu denses (Strasbourg, Nîmes), et des structures urbaines au ***développement relayé en périphérie proche*** où la population est importante dans les noyaux connectés ou satellites (Mulhouse, Le Havre).

5.3.1.3 Analyse de la fragmentation intra-urbaine

Le jeu de données étudié dans cette section correspond au nombre de régions nécessaires pour obtenir trois-quarts de la surface des régions de même morphotype par aire urbaine. Ce deuxième groupe de variables dénombre cette fois-ci des régions d'image et est aussi analysé à l'aide d'une AFC.

Le tableau 5.6 présente les valeurs propres et l'inertie associée pour les 11 axes fournis par l'AFC analysant la fragmentation intra-urbaine. Ici aussi il ne semble pas y avoir de corrélation forte entre les variables ni entre les individus puisque toutes les valeurs propres sont faibles. Le premier axe représente toutefois à lui seul 45% de l'inertie du jeu de données. Le reste de l'inertie se concentre majoritairement sur les 4 premiers axes qui représentent à eux seuls 80 % de l'inertie totale du jeu de données. Pour le reste de l'analyse et des traitements, nous nous intéressons aux cinq premiers axes de cette AFC qui résument des tendances d'organisation basées sur toutes les catégories de morphotypes considérées en entrée.

TABLE 5.6 – Décomposition de l'inertie par axe de l'AFC fragmentation intra-urbaine

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Valeurs propres	0,09	0,04	0,02	0,02	0,01	0,01	0,01	0,00	0,00	0,00	0,00
% d'inertie	44%	17%	12%	8%	6%	4%	3%	2%	2%	2%	1%
% d'inertie cumulée	44%	61%	73%	81%	86%	91%	94%	96%	98%	99%	100%

La figure 5.9 présente le premier plan factoriel de cette AFC. Les facteurs s'y interprètent comme des variables nouvelles opposant des catégories de morphotypes et constituent différentes modalités de la fragmentation des espaces intra-urbains. On constate par exemple sur l'axe 1 l'opposition d'aires urbaines aux nombreux noyaux isolés, comme Besançon, Troyes ou Limoges, à des aires urbaines composées de nombreux centres secondaires, ou de couronnes denses morcelées comme Lille, Marseille-Aix, Toulon ou Nice. L'axe 2 oppose lui aussi des aires urbaines présentant de nombreux centres secondaires comme Strasbourg ou Dunkerque mais cette fois-ci à des aires urbaines présentant de nombreuses zones d'activité périphériques et noyaux satellites. Strasbourg, avec une morphologie particulière où les centres denses de nombreux villages sont détectés comme des centres secondaires, se détache du reste des aires urbaines sur le second axe. Elle contribue de façon non négligeable à sa création.

Globalement l'analyse des contributions et des corrélations entre variables initiales et composantes principales (cf. volume II) indique que ces cinq premiers facteurs permettent de représenter correctement l'ensemble des variables initiales. Dans cette AFC, la catégorie des centres principaux joue un rôle minime dans la différenciation des aires urbaines. Puisque l'échantillon d'étude ne contient pas d'aires urbaines caractérisées par une organisation polycentrique aux centres très denses et de même niveau, cette modalité ne contribue pas fortement à la création des premiers axes.

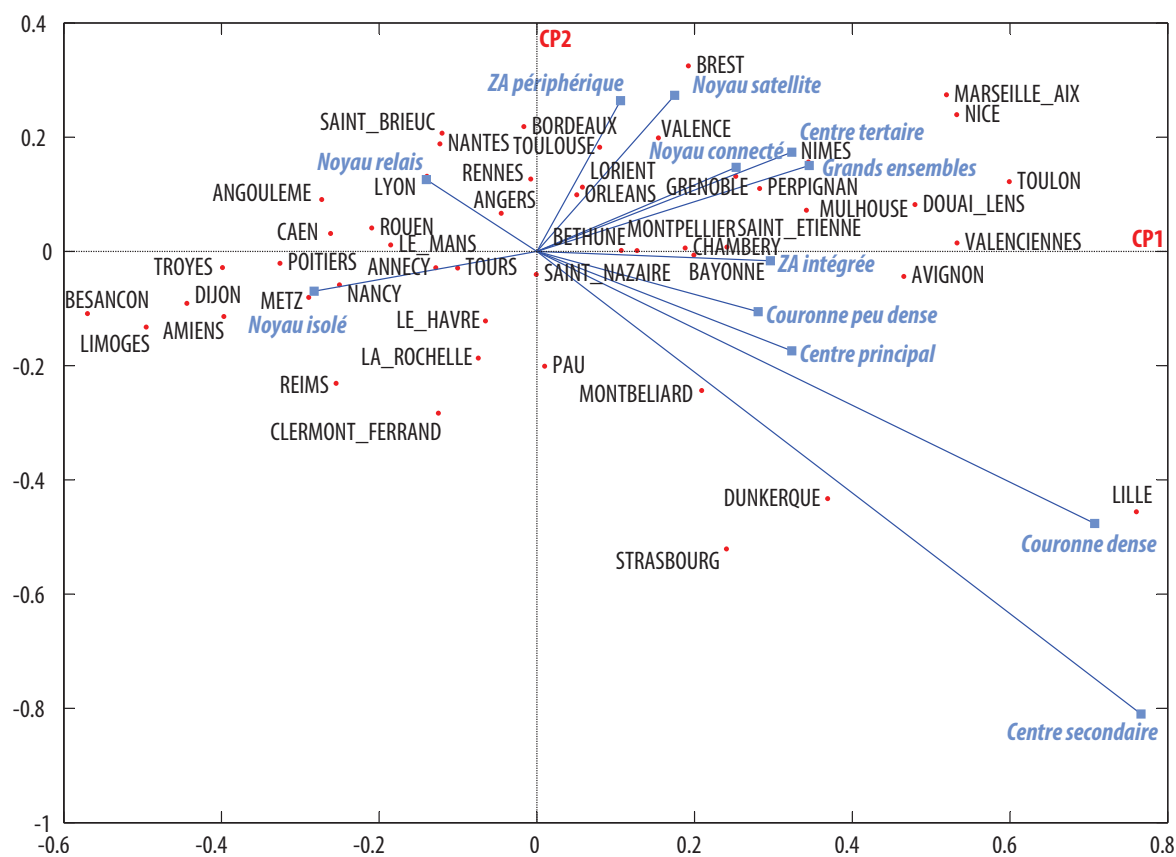


FIGURE 5.9 – Premier plan factoriel de l'AFC fragmentation intra-urbaine

Les grandes tendances de fragmentation des espaces intra-urbains peuvent ainsi être résumées à partir des facteurs suivants :

- C_{frag}^1 : Structures urbaines au **développement périphérique** possédant de nombreux noyaux isolés (Besançon, Limoges) opposées à des structures urbaines à **tendance polycentrique** avec de nombreux centres secondaires et des couronnes denses fragmentées (Marseille-Aix, Lille).
- C_{frag}^2 : Structures urbaines au **développement hiérarchisé et dense** possédant de nombreux centres secondaires et couronnes denses (Strasbourg, Dunkerque) opposées à des structures urbaines au **développement économique périphérique** possédant de nombreuses zones d'activités périphériques et noyaux satellites (Brest, Bordeaux).
- C_{frag}^3 : Structures urbaines au **développement hiérarchisé** avec de nombreux centres tertiaires et couronnes peu denses (Rennes, Perpignan) opposées à des structures urbaines au **développement marqué en périphérie** avec de nombreux noyaux connectés et satellites (Grenoble, Douai-Lens).
- C_{frag}^4 : Structures urbaines au **développement contraint et intégré** remarqué par la présence de nombreux noyaux satellites, grands ensembles et zones d'activités

intégrées (Grenoble, Marseille-Aix) opposées à des structures urbaines à **tendance conurbation** avec de nombreux noyaux connectés et centres principaux (Douai-Lens, Béthune).

C_{frag}^5 : Structures urbaines au **développement économique périphérique et relayé** possédant de nombreuses zones d'activités périphériques et noyaux relais (Bayonne, Brest), opposées à des structures urbaines **hiérarchisées et coalescentes** avec de nombreux centres tertiaires et noyaux connectés (Douai-Lens, Angers).

5.3.1.4 Analyse des relations et du positionnement intra-urbain

Nous nous intéressons maintenant aux résultats de l'AFC appliquée au groupe de variables correspondant au dénombrement des relations d'adjacence entre paires de régions aux morphotypes différents. Rappelons que ces mesures qui informent sur les relations spatiales intra-urbaines et sur le positionnement relatif par rapport au(x) centre(s) des structures urbaines sont données pour toutes les aires urbaines qui ont été divisées en deux espaces : l'espace central associé à l'agglomération principale, et l'espace périphérique complémentaire. Nous parlerons donc de « variables centrales » et de « variables périphériques ».

Le tableau 5.7 présente les valeurs propres et inerties associées à chacun des 17 axes de l'AFC. L'information est fragmentée sur les différents axes, avec un premier axe qui ne contient que 26% de l'inertie totale du jeu de données. Pour conserver plus de 80% de l'inertie totale, il faut considérer au moins les six premiers axes de l'AFC. Les derniers axes sont associés à des valeurs propres et des inerties faibles. La suite de l'analyse concernera donc les six premiers axes qui conservent une part d'inertie suffisante pour représenter les principales caractéristiques d'organisation des aires urbaines.

TABLE 5.7 – Décomposition de l'inertie par axe de l'AFC relations intra-urbaines

	C1	C2	C3	C4	C5	C6	C7	C8	C9
Valeurs propres	0,16	0,13	0,09	0,06	0,04	0,03	0,03	0,02	0,02
% d'inertie	26%	21%	14%	10%	7%	5%	5%	3%	3%
% d'inertie cumulée	26%	46%	60%	70%	76%	82%	86%	90%	92%
	C10	C11	C12	C13	C14	C15	C16	C17	
Valeurs propres	0,01	0,01	0,01	0,01	0,00	0,00	0,00	0,00	
% d'inertie	2%	2%	1%	1%	1%	1%	0%	0%	
% d'inertie cumulée	95%	96%	97%	98%	99%	100%	100%	100%	

La figure 5.10 présente le premier plan factoriel de cette AFC. Le premier axe oppose principalement deux variables centrales : les relations entre centres principaux et secondaires et couronnes peu denses, à la variable périphérique des relations de centres tertiaires à couronnes peu denses. Cet axe contenant la principale part d'information du jeu de données oppose donc en quelque sorte les aires urbaines au développement centré dans l'agglomération principale, comme pour Lille et Dunkerque, aux aires urbaines au développement périphérique marqué par la présence d'agglomérations de plus petite taille (vues par la présence de centres tertiaires) comme pour Rennes, Lorient ou Perpignan.

Le second axe présente la particularité d'exister quasiment exclusivement par la contribution de variables périphériques. Cet axe oppose toutes les structures contenant des agglomérations périphériques développées à celles n'en contenant pas ou peu. Strasbourg, Avignon, Béthune ou Douai-Lens se détachent par leurs coordonnées extrêmes et leur nature nettement plus polycentrique et/ou hiérarchisée que le reste des aires urbaines françaises. Le nom de l'aire urbaine de Douai-Lens indique clairement le polycentrisme de l'organisation constituée de ces deux grandes agglomérations, en plus de petites agglomérations morphologiquement distinctes. A l'opposé, une aire urbaine comme celle d'Avignon se détache par la très forte hiérarchisation des pôles composant

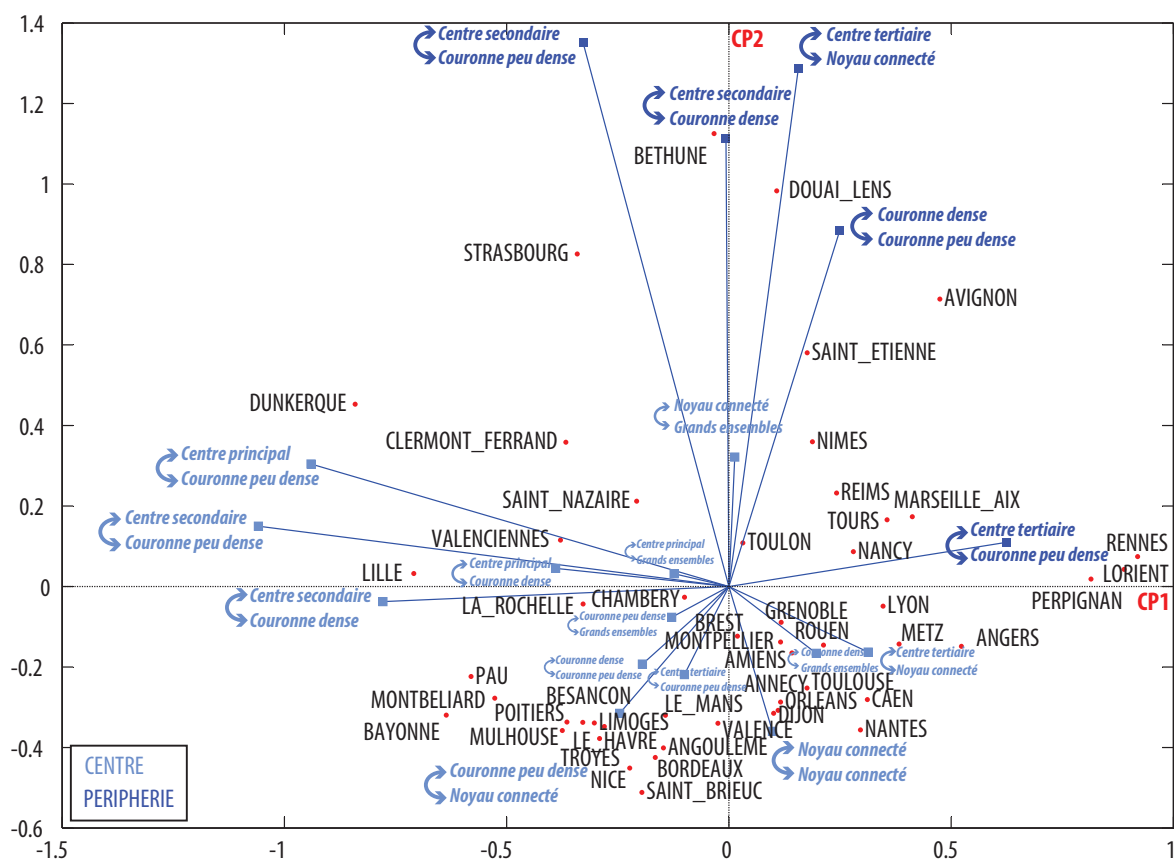


FIGURE 5.10 – Premier plan factoriel de l'AFC relations intra-urbaines

l'aire urbaine. Dans cette structure, on retrouve un nombre important d'agglomérations périphériques de niveau intermédiaire composées de centres secondaires et tertiaires entourés de couronnes denses et peu denses. Une aire urbaine comme Lille, aussi connue pour être polycentrique apparaît cependant très proche de l'origine sur cet axe car son développement diffère d'un point de vue morphologique : elle est l'emblème des structures urbaines correspondant à des conurbations.

Les axes suivants révèlent d'autres tendances d'organisation caractérisant le développement urbain des agglomérations principales et périphériques. Globalement, les variables périphériques semblent être mieux représentées par le groupe des premières composantes principales, ce qui laisse penser que l'information la plus discriminante pour qualifier les structures urbaines sous l'angle des relations spatiales concerne les variations du développement périphérique. Cependant l'hétérogénéité des relations dans les agglomérations centrales constitue elle aussi une information pertinente qui trouve sa part dans la création des premiers axes de l'AFC. En résumé, les six premières composantes principales réorganisent l'information contenue dans les variables initiales en mettant en avant les caractéristiques suivantes de relations intra-urbaines au sein des structures urbaines :

- C_{rel}^1 : Structures urbaines au *développement périphérique hiérarchisé*, marquées par la présence de centres tertiaires entourés de couronnes peu denses (Rennes, Lorient), opposées à des structures urbaines à *agglomération principale aux centres entourés de couronnes peu denses* (Lille, Dunkerque).
- C_{rel}^2 : Structures urbaines aux *agglomérations périphériques développées* et composées de couronnes de différents niveaux (Douai-Lens, Strasbourg), opposées à des structures urbaines à *l'agglomération centrale étalée et axialisée* qui contient de nombreux noyaux connectés entre eux et à la couronne peu dense (Saint-Brieuc, Nice)
- C_{rel}^3 : Structures urbaines *coalescentes* caractérisées par la présence de centres tertiaires et de noyaux connectés dans toutes les agglomérations (Douai-Lens, Béthune), opposées à des structures au *développement périphérique hiérarchisé* et marqué par la présence de centres tertiaires entourés de couronnes peu denses (Rennes, Perpignan).
- C_{rel}^4 : Structures urbaines à *grands ensembles périphériques et agglomérations secondaires* (Strasbourg), opposées à des structures urbaines à *agglomération principale hiérarchisée et dense* intégrant des centres secondaires connectés à des couronnes denses et peu denses (Saint-Nazaire, Valenciennes).
- C_{rel}^5 : Structures urbaines à *grands ensembles positionnés en périphérie* de l'agglomération centrale (Marseille-Aix, Poitiers), opposées à des structures urbaines dont l'agglomération principale possède un *développement relayé et hiérarchisé* et des grands ensembles centraux (Toulouse, Grenoble).

C_{rel}^6 : Structures urbaines à *agglomérations périphériques au développement relayé et grands ensembles centraux* (Douai-Lens, Nantes), opposées à des structures urbaines à *agglomération principale hiérarchisée et peu dense* (Avignon, Montbéliard).

Les différentes composantes principales partielles présentées dans cette section résument donc les caractéristiques qui permettent de qualifier les différences d'organisation observables parmi les 50 aires urbaines étudiées. Synthétisant l'information contenue par les variables initiales dans un espace plus réduit, ces caractéristiques représentent différentes tendances d'organisation qui seront dans la suite analysées simultanément afin de définir les tendances globales qui marquent l'organisation des structures urbaines.

5.3.2 L'analyse globale des caractéristiques d'organisation

Dans cette section nous introduisons les résultats de l'analyse globale des caractéristiques, réalisée par AFMM (cf. section 5.2.1). Les analyses partielles des quatre groupes de variables présentées précédemment sont reprises ici, puisque nous choisissons d'effectuer l'analyse sur les composantes principales partielles plutôt que sur les variables initiales. Les composantes principales partielles les plus représentatives servent d'indicateurs pour représenter différentes caractéristiques d'organisation des aires urbaines :

- la répartition de la population par morphotypes (5 indicateurs)
- la fragmentation des morphotypes (5 indicateurs)
- les relations spatiales entre morphotypes (6 indicateurs)
- la forme des agglomérations principales (2 indicateurs)

L'analyse globale est réalisée sur ces 18 indicateurs plutôt que sur les 44 variables initiales afin de travailler sur un espace plus réduit, mais représentatif de toutes ces variables. La perte d'information liée à la suppression des dernières composantes principales est supposée négligeable compte tenu de leur faible inertie et pouvoir explicatif.

Le tableau 5.8 contient les valeurs propres et l'inertie du jeu de données répartie sur chaque axe de l'AFMM. Le premier axe représente 25% de l'inertie totale du jeu de données, associé à une valeur propre de 2,35. Il n'est que partiellement commun à l'ensemble des groupes¹. Les 8 premiers axes de l'AFMM représentent 88% de l'inertie totale. Les 10 derniers facteurs étant associés à une faible inertie, ils seront ignorés dans la suite de l'analyse pour plus de lisibilité.

1. Lorsque la première valeur propre est proche de $G = 4$ (groupes de variables), le premier facteur peut être considéré comme commun à l'ensemble des groupes (Escoffier et Pagès, 2008)

TABLE 5.8 – Décomposition de l’inertie par axe de l’AFM

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Valeurs propres	2,35	1,50	1,32	0,97	0,84	0,49	0,44	0,36	0,30	0,20
% d’inertie	25%	16%	14%	10%	9%	5%	5%	4%	3%	2%
% d’inertie cumulée	25%	41%	55%	66%	75%	80%	85%	88%	92%	94%
	C11	C12	C13	C14	C15	C16	C17	C18		
Valeurs propres	0,18	0,11	0,10	0,09	0,04	0,03	0,03	0,02		
% d’inertie	2%	1%	1%	1%	0%	0%	0%	0%		
% d’inertie cumulée	96%	97%	98%	99%	99%	100%	100%	100%		

L’interprétation des composantes principales globales repose sur la distinction de caractéristiques dominantes issues des analyses factorielles partielles. Il est donc plus difficile d’interpréter les composantes globales lorsque plusieurs composantes partielles contribuent à part égale et dans des directions égales, puisque ces composantes font appel à un plus grand nombre de tendances d’organisation. Dans ce cas il est intéressant d’observer les projections des individus et des variables, et éventuellement de revenir aux caractéristiques des composantes partielles pour distinguer des informations structurantes dans le mixte des composantes partielles.

La figure 5.11 présente le premier plan factoriel de l’AFMM. Le premier axe regroupe diverses caractéristiques liées à la population, à la fragmentation et aux relations entre régions, et à la forme des agglomérations principales. On y retrouve d’une part des aires urbaines corrélées négativement comme Douai-Lens, Lille ou Strasbourg composées de plusieurs agglomérations distinctes repérées par la composante partielle C_{rel}^2 . Les composantes partielles C_{pop}^1 et C_{frag}^1 permettent d’associer ces agglomérations à la présence de nombreux centres principaux et secondaires qui contiennent une part importante de population. D’autre part on retrouve des aires urbaines comme Troyes, Poitiers ou Amiens, caractérisées par la présence d’une agglomération principale dont le développement peut être qualifié de dense par la composante partielle C_{pop}^2 , et peu contraint par la composante partielle C_{form}^1 . La composante partielle C_{pop}^2 permet aussi d’affirmer que ces aires urbaines possèdent un développement périphérique repéré par la présence de noyaux isolés contenant une part importante de population.

L’axe 2 oppose des aires urbaines corrélées positivement comme Douai-Lens, Béthune ou Valenciennes associées aux composantes partielles C_{rel}^1 et C_{rel}^3 . Ces composantes indiquent un développement périphérique composé d’agglomérations hiérarchisées et relayées par des noyaux connectés. Les aires urbaines corrélées négativement, comme Dunkerque ou Reims, sont associées aux composantes C_{pop}^1 ou C_{pop}^2 qui indiquent une part importante de population dans les centres principaux et secondaires et dans les couronnes denses. Pour des aires urbaines comme Reims ou Besançon, la composante C_{pop}^2 indique aussi la présence d’une part importante en noyaux isolés.

TABLE 5.9 – Contributions des groupes de variables à la création des axes de l'AFMM

	C1	C2	C3	C4	C5	C6	C7	C8
G1 : Population	29%	50%	32%	22%	15%	61%	35%	22%
G2 : Fragmentation	35%	12%	21%	18%	5%	16%	6%	2%
G3 : Relations	22%	37%	41%	59%	11%	14%	55%	73%
G4 : Forme	14%	0%	7%	2%	68%	9%	4%	3%

L'analyse des composantes globales suivantes se poursuit sur le même principe. Pour chacune des composantes principales globales, les différents groupes de variables contribuent de façon plus ou moins équilibrée à la création des axes (cf. tableau 5.9). Les huit premières composantes globales qui permettent de conserver une part importante de l'inertie ne représentent pas à part égale tous les groupes de variables. Le groupe des variables de fragmentation des régions semble moins contribuer à la création de ces composantes. Le détail des contributions des composantes partielles (cf. volume II) indique notamment que les deux dernières composantes partielles de ce groupe n'interviennent que très peu. Les corrélations entre les groupes de variables (cf. tableau 5.10) peuvent expliquer que certaines composantes partielles jouent un rôle moins important.

TABLE 5.10 – Corrélations entre groupes de variables

	G1	G2	G3	G4
G1 : Population	1	0,48	0,35	0,06
G2 : Fragmentation	0,48	1	0,37	0,16
G3 : Relations	0,35	0,37	1	0,09
G4 : Forme	0,06	0,16	0,09	1

Les huit premières composantes principales globales seront retenues pour constituer la typologie de structures urbaines. Elles correspondent aux caractéristiques d'organisation suivantes :

- C_1 : Structures urbaines à la population concentrée (C_{pop}^1), à tendance polycentrique (C_{frag}^1, C_{rel}^2), opposées à des structures urbaines à agglomération au développement peu dense et peu contraint (C_{pop}^2, C_{form}^1) complétées de nombreux noyaux isolés.
- C_2 : Structures urbaines aux agglomérations périphériques hiérarchisées et relayées (C_{rel}^1, C_{rel}^3), opposées à des structures urbaines à agglomérations denses en population (C_{pop}^1) et à périphérie dominée par des noyaux isolés (C_{pop}^2).
- C_3 : Structures urbaines aux centres denses en population (C_{pop}^1) et au développement périphérique à agglomérations hiérarchisées (C_{rel}^1, C_{frag}^3), opposées à des structures urbaines au centre principal développé (C_{pop}^3), marquées par la coalescence des noyaux urbains (C_{rel}^3).

- C_4 : Structures urbaines concentrées (C_{pop}^1) et au développement économique périphérique (C_{frag}^2), opposées à des structures urbaines à agglomérations peu denses (C_{pop}^5) et dont le développement est important à la périphérie (C_{rel}^2, C_{frag}^3).
- C_5 : Structures urbaines dont l'agglomération principale possède un développement non contraint (C_{form}^1) et concentré en population (C_{pop}^1), opposées à des structures urbaines dont l'agglomération principale possède un développement contraint.
- C_6 : Structures urbaines possédant une part importante de population en habitat dispersé (C_{pop}^4), opposées des structures à faible part d'habitat dispersé et à population importante dans les centres principaux et secondaires (C_{pop}^1).
- C_7 : Structures urbaines à grands ensembles périphériques (C_{rel}^5) et noyaux relais développés (C_{pop}^5), opposées à des structures urbaines à grands ensembles périphériques et agglomérations à centres secondaires (C_{rel}^4) et population en habitat dispersé (C_{pop}^4).
- C_8 : Structures urbaines à agglomérations périphériques au développement relayé et grands ensembles centraux (C_{rel}^6), opposées à des structures urbaines aux grands ensembles périphériques et agglomérations secondaires (C_{rel}^4).

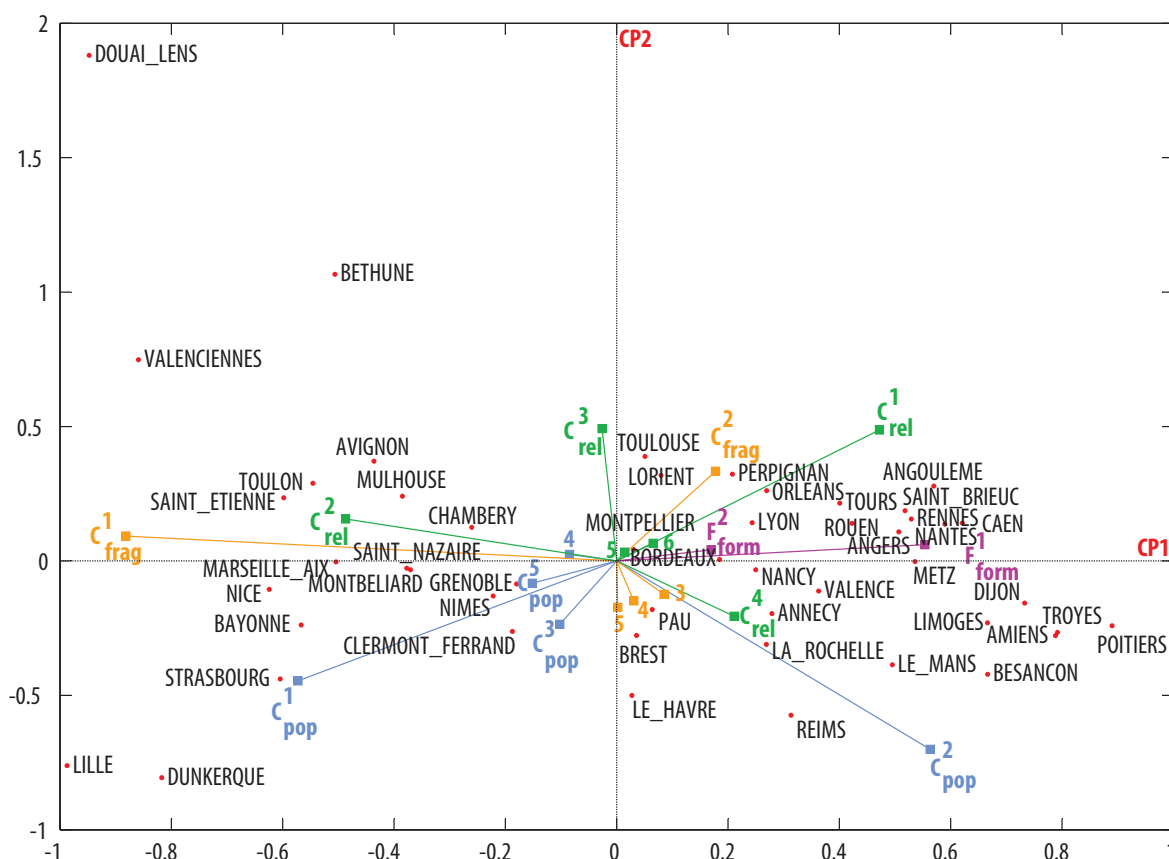


FIGURE 5.11 – Premier plan factoriel de l'AFMM

Ces composantes globales résument donc les caractéristiques principales d'organisation qui différencient les aires urbaines de l'échantillon. Différentes tendances, parfois reliées, concernant la répartition de la population, la fragmentation des régions de même morphotype, les relations spatiales entre régions ou encore la forme de l'agglomération principale sont regroupées sur ces différents axes. A l'issue des analyses partielles et de l'analyse globale, on constate que certaines aires urbaines se distinguent plus des autres quel que soit le groupe de variable considéré. D'autres au contraire semblent se retrouver à des positions assez moyennes sur les différents axes factoriels. A ce stade nous avons identifié plusieurs tendances d'organisation, mais il semble nécessaire de poursuivre l'analyse de l'espace des caractéristiques de taille encore assez importante afin de pouvoir classer les organisations de toutes les aires urbaines de l'échantillon.

5.3.3 Construction de la typologie par CAH

Afin d'obtenir une description des différentes structures urbaines qui puisse dans la suite être résumée graphiquement, nous proposons ici de mener une analyse supplémentaire de ces caractéristiques par classification ascendante hiérarchique (CAH). A partir des composantes principales globales de l'AFMM les plus significatives et représentatives du jeu de données, nous proposons d'identifier des groupes d'aires urbaines caractérisées par des organisations semblables. Cette nouvelle phase de classification servira de base à l'élaboration d'un dernier niveau de modélisation et de représentation sous forme de chorotypes pour chaque groupe d'aires urbaines.

Pour regrouper les aires urbaines selon la similarité de leurs caractéristiques d'organisation spatiale, une CAH est appliquée sur les huit premières composantes principales globales de l'AFMM. La similarité des aires urbaines est mesurée selon le critère de Ward (cf. section 3.1.1.1). Le résultat de ce processus de classification est présenté graphiquement par l'arbre hiérarchique de la figure 5.12 présentant toutes les partitions possibles de $K = 50$ groupes à $K = 1$ groupe, avec une coloration correspondant à une coupe en $K = 11$ groupes.

Le choix d'une partition en 11 groupes a été déterminé à l'aide du dendrogramme figurant la distance entre groupes donnée par le critère de Ward, mais aussi en fonction de l'interprétation des groupes identifiés. La figure 5.13 présente l'évolution du critère par rapport au nombre de groupes. En se basant sur cette figure et sur l'heuristique de coupe du dendrogramme détaillée dans (Rencher, 2002), un palier est identifié à $K = 6$ groupes. Cette heuristique basée sur un seuillage réalisé en comparant les distances entre l'agrégation de deux groupes à la distance moyenne propose une coupe où certains groupes sont très larges. Les plus gros groupes contenant une douzaine d'aires urbaines sont à ce niveau plus difficilement interprétables en termes de caractéristiques d'organisation qui présentent de la variabilité. Sous ce palier à $K = 6$, la progression semble beaucoup plus linéaire et sans palier très important entre deux agrégations. On

repère toutefois un premier palier légèrement plus marqué à $K = 11$, que nous retenons comme un compromis pertinent entre un trop grand nombre de groupes, qui ne refléterait pas les grandes tendances d'organisation spatiale, et un trop petit nombre de groupes, qui donnerait lieu des tendances trop agrégées pour être interprétées.

Les résumés statistiques de ces onze groupes d'aires urbaines selon les composantes principales globales sont donnés dans le tableau 5.11. À partir de ce tableau, il est possible de dégager pour chaque groupe les composantes principales globales qui distinguent les aires urbaines du groupe du reste des aires urbaines. Pour interpréter ces groupes et comprendre les caractéristiques d'organisations qui dominent ces aires urbaines, un élément est déterminant pour isoler les composantes principales explicatives et revenir à leur interprétation en termes de variables initiales : une valeur moyenne éloignée de 0. Si un groupe d'aires urbaines se retrouve à une position centrée sur un axe factoriel, alors l'interprétation en termes de caractéristiques d'organisation est

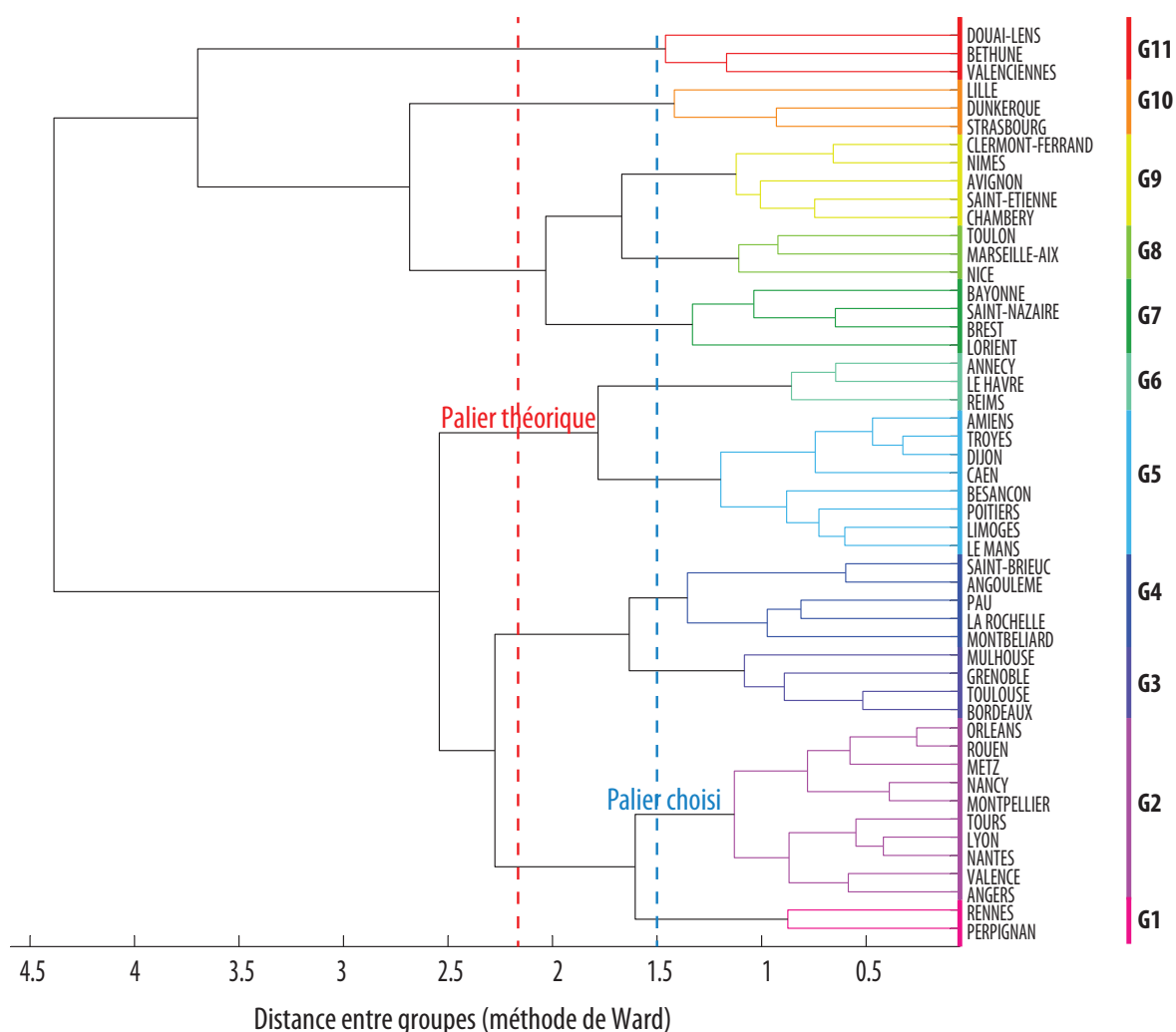


FIGURE 5.12 – Arbre hiérarchique des aires urbaines (partition en 11 groupes)

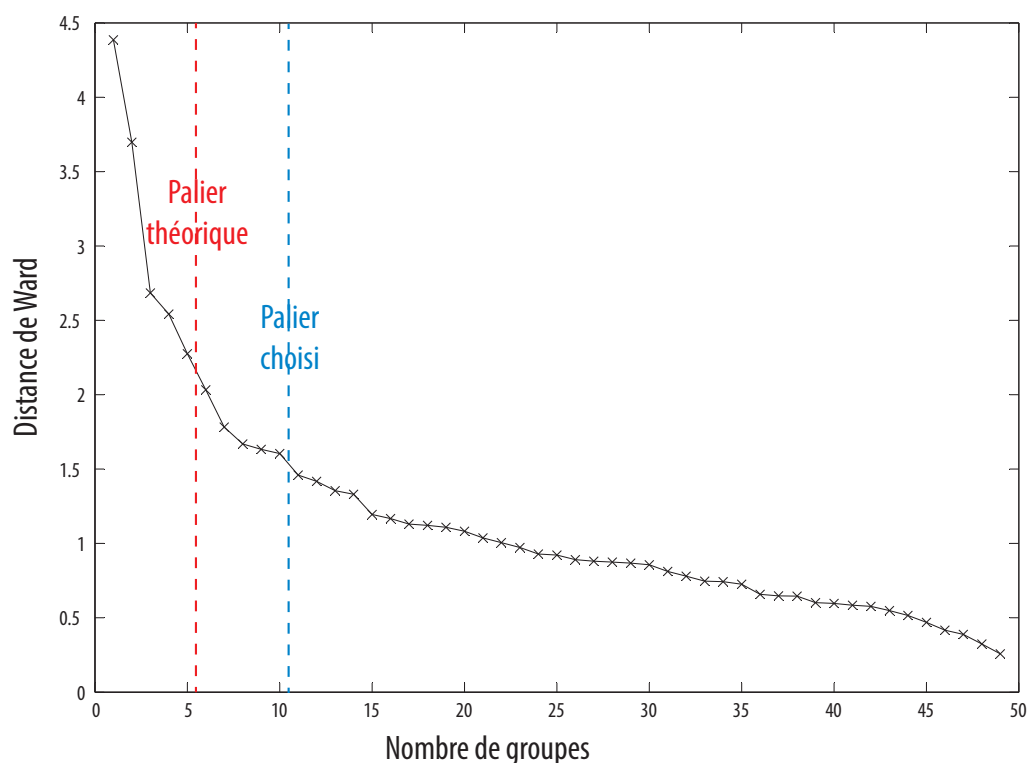


FIGURE 5.13 – Distance de Ward en fonction du nombre de groupes

compromise puisque cette position signifie un profil moyen pour les caractéristiques déterminant l'axe factoriel. Nous nous intéressons donc aux moyennes les plus extrêmes pour chaque groupe et donnons dans ce qui suit, une interprétation des groupes basée sur les composantes principales partielles :

G1 : Structure hiérarchisée

Structure urbaine possédant de nombreuses agglomérations périphériques simples composées de centres tertiaires associés à des couronnes peu denses (C_{rel}^1 , C_{frag}^3).

G2 : Structure macrocéphale hiérarchisée au développement isotrope

Structure urbaine possédant une agglomération principale au développement peu contraint dans toutes les directions (C_{form}^1), composée de centres tertiaires reliés à des noyaux connectés, accompagnée d'agglomérations périphériques composées de centres tertiaires et couronnes peu denses (C_{rel}^1), et relayée par des noyaux relais et satellites et des zones d'activités périphériques (C_{frag}^2).

G3 : Structure étalée et axialisée

Structure urbaine possédant de nombreux noyaux connectés entre eux et à des couronnes peu denses (C_{rel}^2), accompagnés de nombreux noyaux satellites et zones d'activités périphériques (C_{frag}^2).

G4 : Structure périurbaine axialisée

Structure urbaine accueillant une part importante de population en habitat dis-

persé et noyaux isolés, relais, satellites et connectés (C_{pop}^1 , C_{frag}^3), caractérisée par l'absence de relations périphériques (C_{rel}^2).

G5 : Structure périurbaine

Structure urbaine à la population importante dans les noyaux isolés et relais (C_{pop}^1 , C_{pop}^2 , C_{frag}^2), caractérisée par l'absence de relations périphériques (C_{rel}^2).

G6 : Structure périurbaine à agglomération dense

Structure urbaine à la population importante dans les noyaux isolés et relais (C_{pop}^2 , C_{frag}^2), caractérisée par l'absence de relations périphériques (C_{rel}^2), et possédant une agglomération principale accueillant une part importante de population dans les couronnes denses et grands ensembles (C_{pop}^2 , C_{rel}^3 , C_{rel}^6).

G7 : Structure à agglomération au développement contraint, peu dense et hiérarchisé

Structure urbaine dont l'agglomération principale est caractérisée par une forte contrainte de développement en 1/4 de couronne depuis le centre principal (C_{form}^1), par la présence de plusieurs centres secondaires reliés à des couronnes denses et peu denses (C_{rel}^3 , C_{rel}^4), à la population importante dans les couronnes peu denses et en noyaux (C_{pop}^1 , C_{pop}^5).

G8 : Structure polycentrique hiérarchisée dense

Structure urbaine à la population concentrée dans les centres principaux et secondaires, les couronnes denses et les grands ensembles (C_{pop}^1), morphotypes aussi caractérisés par leur fragmentation (C_{frag}^1), et leur répartition aussi bien dans l'agglomération principale que dans des agglomérations périphériques (C_{rel}^1 , C_{rel}^2).

G9 : Structure polycentrique hiérarchisée peu dense

Structure urbaine à population importante en couronnes peu denses et centres secondaires (C_{pop}^5), composée d'agglomérations périphériques développées en centres secondaires, couronnes denses et peu denses (C_{rel}^2).

G10 : Conurbation ou agglomération fragmentée et dense

Structure urbaine à la population concentrée dans les centres principaux et secondaires, les couronnes denses et les grands ensembles (C_{pop}^1 , C_{pop}^2), morphotypes aussi caractérisés par leur fragmentation (C_{frag}^1) et leur association dans l'agglomération principale (C_{rel}^1) accompagnée d'un développement périphérique en agglomérations à centres secondaires et couronnes peu denses (C_{rel}^2).

G11 : Conurbation peu dense

Structure à population importante dans les couronnes peu denses et noyaux connectés et satellites (C_{pop}^1 , C_{pop}^2), à agglomération principale aux centres entourés de couronnes peu denses et de noyaux connectés entre eux (C_{rel}^3 , C_{rel}^4).

L'analyse des groupes à partir des composantes principales partielles révèle que les variables initialement retenues ne jouent pas toutes un rôle fortement discriminant parmi les 50 aires urbaines étudiées. Les composantes les plus discriminantes, qu'elles soient globales ou partielles, sont les premières qui contiennent le plus d'inertie. Même si ce sont souvent les mêmes composantes principales qui aident à discriminer les groupes, tous ont pu être interprétés et associés à des caractéristiques d'organisation différentes.

Ainsi en termes de variables initiales, on constate par exemple que les grands ensembles et leur poids en population jouent un rôle discriminant mais que leur localisation plus ou moins périphérique joue un rôle beaucoup plus minime (visible uniquement sur les dernières composantes globales). La forme des agglomérations principales joue elle aussi un rôle partiel : la directionnalité de développement plus ou moins isotrope depuis le centre principal constitue un fort caractère de différenciation structurelle, mais la nature plus ou moins linéaire, elliptique ou circulaire semble beaucoup moins jouer.

TABLE 5.11 – Profils des groupes de la CAH : moyenne et écart-type des composantes globales

G1			G2		G3		G4	
	Moyenne	Écart-type	Moyenne	Écart-type	Moyenne	Écart-type	Moyenne	Écart-type
C1	0,37	0,23	0,36	0,17	-0,08	0,25	0,21	0,38
C2	0,24	0,12	0,09	0,12	0,14	0,22	-0,01	0,25
C3	-0,69	0,08	-0,07	0,09	0,13	0,31	0,45	0,10
C4	0,37	0,07	-0,01	0,18	-0,43	0,13	-0,16	0,24
C5	0,35	0,23	0,18	0,16	0,02	0,15	-0,11	0,26
C6	-0,35	0,22	-0,01	0,21	0,02	0,25	-0,28	0,22
C7	0,28	0,44	0,04	0,21	-0,20	0,08	0,22	0,27
C8	0,13	0,12	-0,01	0,17	0,09	0,23	-0,11	0,13

G5			G6		G7		G8	
	Moyenne	Écart-type	Moyenne	Écart-type	Moyenne	Écart-type	Moyenne	Écart-type
C1	-0,56	0,06	-0,34	0,17	0,21	0,16	0,71	0,12
C2	0,06	0,20	0,07	0,26	-0,42	0,20	-0,23	0,17
C3	-0,59	0,27	-0,13	0,31	-0,23	0,06	0,32	0,19
C4	-0,26	0,37	0,34	0,22	0,01	0,37	0,08	0,16
C5	-0,08	0,31	-0,09	0,23	-0,20	0,22	-0,18	0,26
C6	0,05	0,30	-0,14	0,16	0,39	0,10	0,10	0,23
C7	0,14	0,23	-0,11	0,16	0,01	0,10	0,13	0,17
C8	-0,25	0,16	-0,09	0,25	0,17	0,13	0,00	0,16

G9			G10		G11	
	Moyenne	Écart-type	Moyenne	Écart-type	Moyenne	Écart-type
C1	-0,21	0,32	-0,80	0,19	-0,77	0,23
C2	-0,06	0,27	-0,67	0,20	1,23	0,58
C3	-0,27	0,37	0,40	0,14	0,66	0,20
C4	0,11	0,27	0,28	0,57	0,24	0,41
C5	-0,75	0,33	0,24	0,32	-0,07	0,29
C6	-0,23	0,19	-0,06	0,16	0,08	0,32
C7	-0,03	0,10	-0,09	0,27	0,22	0,26
C8	0,34	0,09	-0,01	0,31	0,17	0,08

L'interprétation des groupes réalisée à partir des composantes principales globales et partielles révèle donc que différents groupes d'aires urbaines, qu'ils soient petits ou grands, se distinguent très bien les uns des autres grâce à différentes caractéristiques de leur organisation spatiale. L'approche hiérarchique nous a permis d'identifier des grandes tendances d'organisation qui différencient progressivement les aires urbaines. On constate ainsi qu'une des tendances de différenciation majeure, qui se retrouve au plus au niveau de l'arbre hiérarchique, concerne le caractère polycentrique ou monocentrique des aires urbaines. Sur l'échantillon d'étude, la majorité des aires urbaines se retrouve dans des groupes à caractère monocentrique, plus ou moins relayé par des centralités tertiaires. Les aires urbaines à tendance polycentrique constituent des cas moins fréquents, et d'autant plus rares quand les centres présents constituent un système peu hiérarchisé. La hiérarchisation des lieux est d'ailleurs elle aussi une caractéristique qui intervient comme une source importante de variations présentes dans les sommets de l'arbre hiérarchique.

Le caractère de connexité des formes joue aussi un rôle majeur dans la différenciation des aires urbaines. Les variables initiales analysées par regroupement hiérarchique permettent de différencier des formes de connexions et de distinguer les organisations composées de plusieurs agglomérations, de celles composées d'une unique agglomération qui peut prendre une forme plus ou moins coalescente et axialisée. Ces variables permettent aussi d'observer d'autres tendances majeures comme la densité plus ou moins importante des agglomérations, ou le caractère plus ou moins périurbain des aires urbaines observé par la présence des noyaux d'urbanisation qui complètent la composition des organisations.

L'analyse factorielle a donc permis de dégager différentes tendances marquant l'organisation des aires urbaines à partir de vecteurs de caractéristiques composés initialement de 44 variables décrivant la forme des agglomérations, la répartition de la population, la fragmentation et les relations entre régions de morphotypes différents. Ces tendances représentées par les composantes principales globales d'une analyse factorielle de données mixtes ont été exploitées et spatialisées en réalisant une CAH des aires urbaines. Le résultat de cette CAH est l'identification de groupes d'aires urbaines aux caractéristiques d'organisation globalement similaires qui définissent différents modèles d'organisation. Ces différences d'organisation constituent des variations à un niveau intermédiaire entre le modèle chorématique générique et les modèles particuliers que nous avons introduits au chapitre 4. Nous verrons dans la suite que ces tendances d'organisation, et les modèles qu'elles définissent, représentent des caractéristiques importantes pour appréhender des enjeux d'aménagement et pour comprendre le contexte de développement et de fonctionnement de chaque aire urbaine.

5.4 Chorotypes : la spatialisation des enjeux d'aménagement

Dans cette dernière section nous revenons vers la démarche de modélisation chorématique introduite dans le chapitre précédent afin de poursuivre la démarche de spatialisation et d'interprétation graphique des résultats de l'analyse factorielle et de la CAH réalisée précédemment. L'objectif est ici de proposer une version graphique sous forme de chorotypes des grandes catégories d'organisation urbaine identifiées à l'aide de mesures quantitatives. Ces chorotypes complètent le travail de description des aires urbaines entrepris dans cette étude, en proposant un résumé graphique des organisations urbaines à la dernière échelle d'étude macro-géographique.

Nous introduisons donc dans un premier temps les chorotypes qui reprennent les grandes tendances d'organisation correspondant aux 11 groupes d'aires urbaines issus de la CAH. Cette représentation nous permettra ensuite d'introduire une mise en perspective des caractéristiques d'organisation face à différents enjeux d'aménagement.

5.4.1 Les modèles en chorotypes

A l'issue du processus de classification et d'interprétation des caractéristiques de chaque groupe, une visualisation est réalisée sous forme de chorotypes présentés en figure 5.14. Cette représentation permet d'associer à chaque groupe une schématisation des principales caractéristiques d'organisation en reprenant la charte graphique du modèle chorématique générique et des modèles particuliers du chapitre 4.

Pour créer ces représentations graphiques nous reprenons la nomenclature des morphotypes et nous faisons figurer les catégories de formes qui constituent une source de particularité dans l'organisation des groupes d'aires urbaines. La non représentation de certains morphotypes ne signifie pas une absence de ces catégories de formes dans les aires urbaines, mais plutôt que leur présence ne joue pas un rôle structurant important.

La figure qui reprend les ramifications de l'arbre hiérarchique de la CAH présente les groupes d'aires urbaines selon les proximités et les variations dans les chorèmes qu'ils mettent en jeu. La première ligne correspond aux chorotypes représentant des organisations polycentriques, où les variations concernent le chorème de contacts et mise à distance qui crée des structures sous forme de conurbation ou de réseau d'agglomérations. Elles concernent aussi le chorème de dynamique territoriale qui est associé à différentes formes de gradients décroissants, l'une traduisant des structures de faible densité et l'autre traduisant des structures à forte densité.

On retrouve ainsi deux groupes d'aires urbaines, toutes du Nord-Est de la France, représentant des organisations de type conurbation dense ou peu dense marquées par la

présence de multiples centres à haut niveau de densité au sein d'une même agglomération. Les aires urbaines peu denses de l'ancienne bassin minier (Douai-Lens, Béthune, Valenciennes) possèdent une organisation très semblable et distinguable du reste des aires urbaines françaises, où l'ancien habitat minier crée de grandes régions coalescentes très peu denses et guidées par le réseau routier. Le groupe des conurbations denses de Lille, Strasbourg et Dunkerque est un peu plus particulier car si Lille, associée à Roubaix et Tourcoing, représente un exemple classique de conurbation, l'organisation de Strasbourg et Dunkerque résulte de dynamiques légèrement différentes. Les multiples centres observables sur ces deux aires urbaines sont plus associables à une fragmentation des faubourgs anciens aux constructions et à la population très denses, qui n'ont pas nécessairement été associés à des pôles aussi indépendants et structurants que dans le cas de Lille ou des conurbations peu denses.

Sur cette ligne de structures polycentriques, on retrouve aussi le groupe particulier des quelques aires urbaines littorales (Bayonne, Saint-Nazaire, Lorient et Brest) ayant une organisation polycentrique hiérarchisée et peu dense, qui se distinguent par la présence d'une agglomération principale au développement fortement contraint en quart ou demi-couronne.

La deuxième ligne regroupe des chorotypes représentant les organisations monocentriques et fortement dominées par une agglomération principale. Les variations observées se font par exemple sur le chorème de la hiérarchisation qui prend forme selon le nombre de niveaux de centralité présents, sachant que les centres principaux, secondaires et tertiaires ainsi que les noyaux relais, satellites et isolés peuvent être interprétés comme les échelons d'une hiérarchie. Rennes et Perpignan se distinguent dans un groupe marqué par une très forte présence de petites agglomérations périphériques composées de centres tertiaires et couronnes peu denses. Ce groupe est très proche du groupe d'aires urbaines contenant des agglomérations macrocéphales et hiérarchisées comme Lyon, Tours, Montpellier ou Nantes. Le chorotype associé à ce groupe est intéressant puisqu'il est le plus proche du modèle générique d'organisation introduit dans le chapitre 4. Mise à part la sous-représentation des agglomérations à centres secondaires, on retrouve pour ces aires urbaines tous les morphotypes et donc les caractéristiques de développement supposées être celles d'une forme de développement métropolitain.

Le chorotype suivant est plus particulier puisqu'il met en avant le caractère axialisé des structures, observé par la mise en relation de multiples régions de type noyau connecté au sein de la seule agglomération principale. Ce chorotype traduit l'impact plus fort du chorème de contact et mise à distance, et de la structuration par les réseaux de transport. Pour les aires urbaines comme Toulouse ou Bordeaux, ce chorotype traduit clairement un développement radioconcentrique guidé par les réseaux. Mais dans le cas de Grenoble, le développement axialisé résulte surtout de contraintes topographiques autour de l'agglomération.

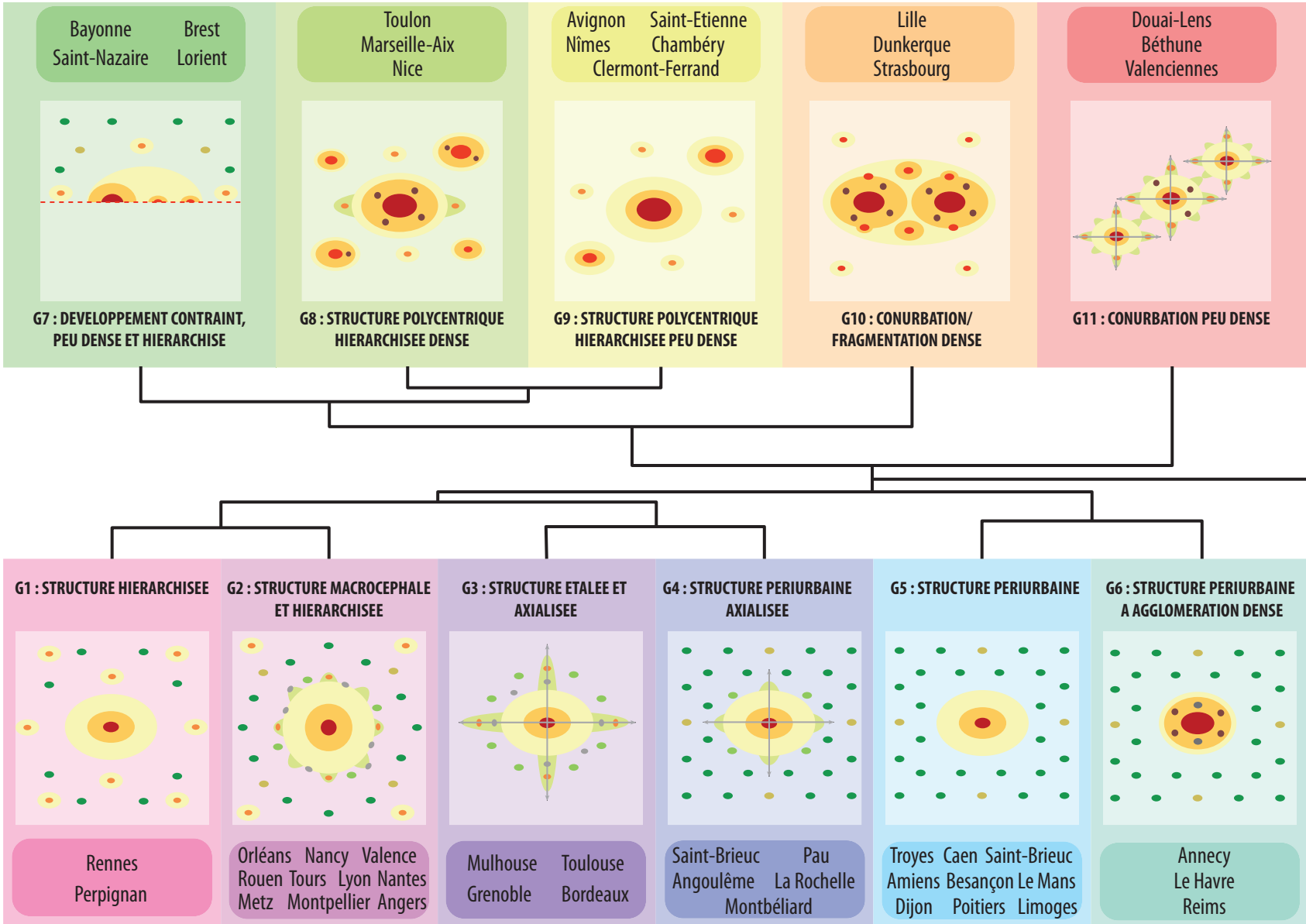


FIGURE 5.14 – Chorotypes de structure urbaine et organisation hiérarchique interclasse correspondant à la partie supérieure du dendrogramme

Les trois derniers chorotypes renvoient à la fragmentation du maillage urbain et à la dynamique territoriale fortement marquée par la périurbanisation des lieux de résidence. Les différences entre les trois groupes concernent le développement de l'agglomération principale qui dans un cas est marquée par des tendances d'étalement et de coalescence des régions, dans un autre est marquée par une densification importante, et dans le dernier constitue un intermédiaire entre les deux. On retrouve de nombreuses aires urbaines de l'échantillon d'étude marquées par ces caractéristiques d'organisation.

Ces différentes représentations chorématiques constituent donc des modèles intermédiaires résumant les différents modèles d'organisation particuliers qui peuvent être observés à partir des données sur grille régulière et des classifications des morphotypes. Il est important de noter que dans cette représentation chorématique, la forme de base, c'est-à-dire l'ellipse, représente volontairement une surface de travail qui n'est pas « la plus neutre possible » (Brunet, 1986). Ce choix permet d'une part de représenter la forme banale observée pour la majorité des agglomérations et d'autre part de mettre en avant la singularité du développement isotrope du groupe 2. Nous proposons maintenant d'analyser ces différents chorotypes face aux enjeux d'aménagement qui peuvent être soulevés de façon générale et qui se retrouvent dans les différentes aires urbaines étudiées.

5.4.2 Modèles urbains et enjeux d'aménagement

Au-delà du simple retour aux processus de spatialisation des organisations urbaines, ces chorotypes peuvent révéler des enjeux de développement urbain qui prennent tout leur sens lorsqu'ils sont remis dans le contexte de l'étude des politiques d'aménagement des territoires. Pour mettre en perspective les différents chorotypes, nous nous servons des clés d'interprétation des principaux enjeux d'aménagement tirées de documents d'aménagement tels que le SCoT (Schéma de Cohérence Territoriale) ou le PLU (Plan Local d'Urbanisme). Les orientations de planification fixées dans ces documents recouvrent plusieurs thématiques de l'organisation des espaces urbains qui sont représentées par nos classifications en tissus urbains et en morphotypes, et en définitive par la typologie en chorotypes.

Il est important de noter que ces documents sont élaborés par des groupements de communes correspondant à des EPCI (Etablissement Public de Coopération Intercommunale) qui ont rarement une étendue coïncidant avec le périmètre que nous avons choisi pour construire les représentations, à savoir celui des aires urbaines. Le périmètre de l'aire urbaine englobe dans certains cas un ou plusieurs EPCI, mais dans d'autres il est nécessaire d'accoler plusieurs aires urbaines pour reconstituer le territoire de l'EPCI. La confrontation des résultats de la typologie des structures urbaines aux choix et vœux en matière d'aménagement doit donc être faite en tenant compte de ces différences d'étendue spatiale.

5.4.2.1 L'intensité urbaine

Un des volets présents dans de nombreux projets urbains concerne la promotion de la « ville intense », à savoir la ville dense et mixte desservie par des transports collectifs, équipements et services de proximité. Cette vision de la ville renvoie aux questions de densité urbaine, ou d'intensité urbaine par opposition à l'extension urbaine. L'intensité urbaine est promue dans divers SCoT comme un moyen de lutte contre la consommation d'espace, problème ciblé par exemple dans le SCoT de la métropole de Montpellier² ou de Nantes Saint-Nazaire³. Le mot d'ordre est alors celui du « renouvellement urbain », de « la ville sur elle-même » ou de « la ville dans la ville ». Les mesures concernent alors des taux de nouvelles constructions qui doivent se faire dans le tissu urbain existant, l'imposition de densité minimale ou encore par le renforcement de la densité le long d'axes de communication spécifiques, notamment les lignes de transport en commun.

Il est intéressant de confronter ce désir d'intensité urbaine aux différents chorotypes identifiés puisque la densité de population et de surface bâtie est une composante intégrée à tous les morphotypes composant les organisations urbaines. Ce principe de développement urbain n'aura en effet pas la même latitude dans les organisations les plus denses représentées par des chorotypes tels que les structures polycentriques, les conurbations denses, ou les autres chorotypes à agglomération dense. Même dans des métropoles telles que Marseille⁴, Toulon⁵ ou Lille⁶ l'intensité urbaine est préconisée dans les SCoT. Cette recherche de densité prend cependant une dimension particulière dans les territoires les moins denses correspondant au chorotype G11 des conurbations peu denses. Si les SCoT des métropoles de Douai⁷ ou de Lens⁸ mentionnent la densité, c'est en premier lieu par un constat de manque d'attractivité pour ces territoires qui ressortent affaiblis d'une phase de reconversion industrielle. Dans cette aire urbaine à la périphérie de Lille, la densité est recherchée pour signifier l'attractivité des territoires, mais aussi pour lutter contre la périurbanisation lilloise qui déborde sur ces territoires.

5.4.2.2 Le transport et les mobilités

Comme de nombreuses études attachées à la description des formes urbaines l'ont déjà montré, cette recherche de la ville intense est à relier à la mise en place d'une politique de transport efficace. Les objectifs retrouvés dans divers SCoT en matière de transport sont assez similaires quel que soit le type d'organisation urbaine rencontré. Ils concernent principalement la favorisation de modes de déplacement alternatifs à la

2. <http://www.montpellier-agglo.com/connaître-compétences/aménagement-de-l-espace-communautaire>

3. http://www.nantessaintnazaire.fr/web/menu_le_scot_de_la_metropole.do

4. <http://www.marseille-provence.com/thematiques/habitat-urbanisme-politique-de-la-ville/un-schema-de-coherence-territoriale.html>

5. <http://www.scot-pm.com/index.php?id=137>

6. <http://www.scot-lille-metropole.org/>

7. <http://www.scot-douais.org/>

8. <https://www.scot-llhc.fr/>

voiture particulière et la mise en place ou le renforcement de pôles intermodaux.

Or la mise en place de réseaux de transports collectifs performants est conditionnée par la demande et donc par la densité suffisante de population autour des points d'accès à ces réseaux. Si la desserte de l'agglomération principale peut se faire à partir d'une solution de type transport en commun en site propre selon un principe de « ville de proximité », elle doit se faire selon d'autres principes dans les zones périphériques. Un développement articulé autour des réseaux ferrés est notamment privilégié mais aussi un développement articulé autour du réseau routier pour organiser les points de rabattement ou de contournement des zones les plus congestionnées. De plus, si la densité résidentielle conditionne l'origine plus ou moins dispersée des flux de déplacements, la polarisation des espaces urbains et la répartition des activités économiques conditionnent une grande partie des destinations de ces flux. Les zones de centralité ainsi que les zones d'activités constitueront d'importants points d'arrivée des déplacements. Les chorotypes de structure urbaine identifiés se révèlent d'intérêt pour l'élaboration de politiques de transport en spatialisant les lieux résidentiels, sources de déplacement, et leur densité ; en spatialisant la répartition des points potentiels d'arrivée ; mais aussi en spatialisant la densité des lieux à traverser (ou à éviter) pour rallier les points de départ et d'arrivée.

5.4.2.3 La multipolarité

Les questions sur la multipolarité ou les multipolarités au pluriel, tant ce mot renvoie à des situations différentes, constituent un enjeu d'aménagement identifié par de nombreuses métropoles. La mise en avant de ce concept traduit la volonté de mettre en œuvre une organisation équilibrée du territoire. Cette forme urbaine préconisée dans les plans d'urbanisme est vue par les aménageurs comme le projet urbain qui assurera une meilleure mixité fonctionnelle (redistribution des activités centrales en de multiples centres) et des distances de déplacements plus courtes, bien que ces relations ne soient pas systématiques vérifiées (Schwanen et al., 2004).

Dans l'ouvrage de Grosjean (2009) qui compare les formes urbaines et plans d'aménagement des métropoles de Lille, Bordeaux⁹, Montpellier et Nantes Saint-Nazaire, l'auteure note que la multipolarité promue par ces plans d'aménagement recouvre en pratique des réalités très différentes que nous retrouvons dans notre représentation et typologie d'aires urbaines. La métropole de Lille (qui coïncide avec l'aire urbaine) représente le cas d'école français de conurbation où la multipolarité souhaitée correspond à un fonctionnement polycentrique associé à un certain équilibre entre pôles. Les polarités y sont cependant de plusieurs niveaux identifiés dans le SCoT et dans notre classification : les trois têtes de pont Lille, Roubaix et Tourcoing identifiées en centre principal, les polarités secondaires de l'espace aggloméré comme la ville nouvelle

9. <http://www.sysdau.fr/>

de Villeneuve d'Ascq, ou encore les polarités périphériques hors de l'espace aggloméré identifiées pour certaines en centre tertiaire¹⁰.

A l'opposé, la multipolarisation mise en avant par la métropole de Bordeaux diffère par ses objectifs et sa nature observée : un unique pôle fort, le centre principal, est identifié au cœur de Bordeaux ; les polarités secondaires identifiées dans le SCoT sont par comparaison des polarités beaucoup moins marquées et représentées dans notre cas au mieux par des centralités de niveau tertiaire (Merignac ou Pessac), par le morphotype de noyau connecté (Ambarès-et-Lagrave) ou même par une intégration totale aux couronnes denses et peu denses (Talence, Bègue), ce qui ne signifie pas une absence totale de ces centralités, mais l'absence d'un caractère morphologique discriminant ces pôles en termes de population ou de bâtiments. La multipolarité souhaitée par cette métropole ne correspond pas à une vision polycentrique, mais plutôt à une vision macrocéphale telle que nous l'identifions dans la typologie. Le centre de l'agglomération centrale au développement très important se retrouve relayé par des centralités tertiaires correspondant à des lieux anciennement indépendants morphologiquement.

Fruit d'une politique d'aménagement mise en place depuis les années 80, le cas de Rennes¹¹ constitue un troisième archétype de la multipolarité cette fois représentée par le chorotype des structures hiérarchisées proche des configurations christalleriennes. Dans ces aires urbaines, de multiples lieux de centralité sont observables et distinguables morphologiquement mais ils restent principalement soumis à un pôle central au développement beaucoup plus conséquent. Agglomération compacte et développement urbain appuyé par les bourgs périphériques originels, ce type d'organisation renvoie aux systèmes hiérarchisés de Christaller en termes de répartition des services mais constitue aussi un atout de développement urbain mis en avant en termes de gestion de l'étalement urbain et de préservation des ressources paysagères, naturelles et agricoles.

5.4.2.4 La localisation des activités économiques

D'autres volets identifiés dans les SCoT concernent par exemple la localisation du développement économique. Enjeu caractéristique des villes moyennes (DIACT, 2007), la répartition spatiale des activités économiques et plus particulièrement le maintien des activités en centre-ville constitue un point de développement pris en compte dans divers plans d'aménagement. Nantes, structure macrocéphale isotrope caractérisée par un grand nombre de zones d'activités périphériques préconise par exemple de mieux gérer la répartition spatiale des activités en interdisant l'implantation de nouvelles zones d'activités périphériques. Cette préoccupation se retrouve aussi dans le SCoT de Gre-

10. le rôle moteur de ces polarités identifiées dans le schéma directeur de 2002 est remis en cause pour certaines dans le diagnostic territorial de 2010 du SCoT en cours d'élaboration

11. <http://www.paysderennes.fr/-Notre-SCoT-.html>

noble¹², structure étalée et axialisée aux zones d'activités périphériques pourtant moins nombreuses. La volonté exprimée n'est pas d'enrayer totalement le développement de zones d'activités périphériques mais plutôt de filtrer le type d'activités économiques qui s'y implantent pour assurer un maintien des centres villes et de leurs services de proximité.

Conclusion

Ce chapitre a permis d'introduire un dernier niveau d'analyse et de représentation des organisations urbaines caractérisées ici à l'échelle de la macro-forme. Cette représentation directement en lien avec les classifications en tissus urbains et en morphotypes est basée sur l'étude de différentes mesures de caractéristiques d'organisation. Ces vecteurs de caractéristiques ont été traités à partir d'analyses factorielles multiples adaptées au traitement de données mixtes. Ces analyses ont permis de produire un ensemble résumant les tendances d'organisation spatiale en termes de répartition de la population, de fragmentation et de relations entre morphotypes, et de forme du développement de l'agglomération principale.

A l'issue de ces analyses factorielles, il a été possible de construire une typologie des structures urbaines révélant différents types d'organisation morphologique observables sur le territoire français. Cette typologie offre une source d'information associée aux représentations sur les grilles régulières des structures urbaines qui présente l'intérêt de pouvoir servir de support d'étude à de nombreuses questions d'aménagement reliées au développement morphologique des agglomérations. La diversité des configurations morphologiques s'est révélée être un enjeu d'intérêt pour aborder les questions de consommation d'espace et d'intensité urbaine, les questions de politique de transport et de gestion de la composition territoriale.

Dans la mesure où la morphologie représente un des angles d'étude pour sonder la complexité urbaine, un examen du lien entre diversité morphologique et diversité fonctionnelle constituerait un enrichissement pertinent pour développer et enrichir la typologie des structures urbaines.

12. http://www.region-grenoble.org/index.php?option=com_content&view=article&id=53&Itemid=253

Conclusion générale

Synthèse des travaux

L'objectif général de cette recherche a été de proposer une représentation de la morphologie du bâti observée dans le périmètre des aires urbaines françaises. Deux enjeux ont guidé ce travail de modélisation :

- Nous avons d'une part cherché à faire ressortir les éléments les plus robustes d'une forme urbaine, principalement décrite par sa population résidente et son bâti caractérisé par des fonctions, de façon à construire des modèle de structure.
- Nous avons d'autre part cherché à mettre au point des procédures de classification automatique de données empiriques disponibles à l'échelle nationale, avec le minimum d'intervention d'un opérateur.

Les données de population nouvellement diffusées par l'INSEE sur des grilles régulières à mailles de 200 mètres de côté ont permis d'utiliser un nouveau référentiel spatial pertinent pour reconstituer ces organisations. A partir de mesures de densité de population complétées par des mesures de densité et de diversité relatives aux surfaces bâties, nous avons proposé une caractérisation de la morphologie des aires urbaines à trois échelles d'études : l'échelle du pixel, l'échelle de la région d'image et l'échelle de la macroforme. Ces travaux synthétisent donc une réflexion théorique appuyée par un important volet expérimental invoquant diverses méthodes de classification de données spatialisées appliquées à un échantillon varié d'aires urbaines.

Des modèles théoriques et graphiques ...

Nous avons cherché à identifier des éléments théoriques qui soient adaptés à la représentation des formes urbaines métropolitaines que nous associons à un modèle de développement urbain caractéristique du 21ème siècle. Le parti-pris de ce travail a été de se baser sur une approche morphologique des espaces urbains considérés selon les

relations qu'entretiennent des lieux différenciés par des caractéristiques de forme, de fonction ou de hiérarchisation (pour des lieux marqués par différents degrés de centralité). En prenant une démarche de représentation basée sur divers travaux modélisation urbaine, mais aussi en se basant sur la vaste littérature sur les liens entre formes urbaines et pratiques de déplacement, nous avons fait l'hypothèse que ces configurations morphologiques représentaient un déterminant du fonctionnement des organisations urbaines. En tant que support différencié des hommes et des activités, la structure morphologique et fonctionnelle est un cadre pour les mouvements urbains, qu'ils soient inscrits dans des temporalités courtes (les déplacements quotidiens) ou longues (la morphogénèse).

Pour guider l'identification des caractéristiques morphologiques qui nous semblent indispensables pour comprendre les processus marquant l'organisation des espaces urbains, nous avons travaillé à différentes échelles sur les grilles régulières. Ce travail nous a permis de définir deux nomenclatures à appliquer aux données de densités sur grille régulière : une nomenclature en tissus urbains et une nomenclature en morphotypes identifiés à partir d'un modèle chorématique général des formes métropolitaines. Le recours aux données de densité sur grille régulière nous a offert le moyen de tester ces éléments de modélisation théorique et d'observer différentes classes de structures morphologiques et fonctionnelles, plus particulières et définies comme des chorotypes.

... spatialisés par la classification de données

Le cadre quantitatif d'analyse de la morphologie du bâti basé sur la densité de population résidente et de surface bâtie sur des grilles régulières offre le moyen d'analyser la forme urbaine comme une image à laquelle on peut appliquer diverses méthodes issues du traitement d'image et du raisonnement spatial. Cette image a servi de support à l'identification des tissus urbains et des morphotypes conçus comme les régions d'image homogènes issues de classifications.

Une des contributions importantes de ce travail de thèse a été de proposer une méthode de classification des tissus urbains applicable quelle que soit l'aire urbaine étudiée, et sans connaissance spécifique de l'espace à analyser. Un algorithme de classification supervisée par transfert d'apprentissage a été mis au point pour traiter de façon générale toutes sortes de données spatialisées. Le formalisme des champs de Markov discrets cachés a été employé pour sa capacité à gérer les données présentant des dépendances statistiques et spatiales.

La solution proposée a été employée pour obtenir les classifications en tissus urbains de 50 aires urbaines différentes, à partir de connaissances acquises sur seulement 4 aires urbaines. Cette approche nous a donc permis de construire économiquement une base de données homogénéisée des tissus urbains sur toutes les aires urbaines. Certains éléments de la méthode nécessitent encore d'être approfondis, notamment la

sélection automatique du modèle de transfert le plus pertinent pour représenter les liens de distribution statistique variables selon les paires d'aires urbaines considérées. Ce premier niveau de description des organisations morphologiques offre un support de connaissance qui peut être utilisé pour différentes problématiques d'aménagement. Ces classifications en tissus urbains offrent une source de données en partie comparable à ce qui peut être extrait des MOS. La méthode que nous proposons présente l'avantage d'être plus automatisable que les méthodes de photo-interprétation utilisées actuellement pour construire les MOS.

La deuxième contribution de ce travail de thèse a donc été de proposer une solution formalisée dans le cadre du raisonnement spatial qualitatif pour inférer le passage d'un ensemble d'éléments de modélisation conceptuel à des modèles particuliers associés à des données observées et des configurations urbaines particulières. En utilisant cette première base de données des tissus urbains nous avons pu proposer une représentation enrichie de la morphologie des aires urbaines, basée sur la classification des tissus urbains en différents morphotypes de développement urbain. Ce travail de description des aires urbaines se présente comme une transposition directe des modèles chorématiques aux particularités d'organisation morphologique de chaque aire urbaine. En définissant un cadre de représentation et de raisonnement spatial sur les images classées en tissus urbains, nous avons pu proposer un ensemble de règles de décision basées sur des caractéristiques de forme et surtout des relations spatiales topologiques permettant d'inférer ces morphotypes de développement.

La solution proposée a pu être appliquée de façon systématique à toutes les aires urbaines de l'échantillon d'étude, ce qui a abouti à la production d'une seconde base de données homogénéisée décrivant les structures urbaines. A partir de ces représentations comparables sur chaque aire urbaine, nous avons pu définir un dernier niveau d'analyse et de description des configurations morphologiques à l'échelle cette fois-ci du système urbain. Ce dernier niveau d'étude correspond à une analyse plus globale des structures urbaines qui ont été classées en grands types d'organisation, distingués selon les caractéristiques les plus marquantes de leur structure. En utilisant une représentation quantitative et en définissant un espace de caractéristiques analysées à partir de méthodes d'analyse factorielle de données mixtes, nous avons pu mettre en avant les différentes spatialisations des caractéristiques de centralité, de hiérarchisation, de forme de développement et de relation entre ces formes. Toutes ces caractéristiques, nous l'avons vu, sont déterminantes pour mettre en place des politiques d'aménagement et de développement durable, aussi bien vis-à-vis des pratiques de déplacement que de l'orientation des formes de développement.

Perspectives

Plusieurs perspectives peuvent être dégagées pour approfondir ce travail de thèse et le compléter. La première et la plus évidente serait de poursuivre la classification d'autres aires urbaines françaises pour obtenir une description complète du territoire français. La difficulté attendue est sans doute la nécessité de définir de nouveaux prototypes d'aires urbaines permettant de modéliser les plus petites, ou au contraire de modéliser la plus grande, Paris, qui se distingue très fortement du reste des aires urbaines françaises.

Nous distinguons les perspectives suivantes selon les enjeux géographiques et théoriques ou méthodologiques qu'elles soulèvent.

Perspectives géographiques

Si ce travail de thèse est focalisé sur la composante morphologique des aires urbaines, une perspective intéressante serait de confronter ces organisations morphologiques à des organisations fonctionnelles conçues à partir de flux de mobilité. Les liens entre ces deux types d'organisation ont souvent été évoqués. Une suite intéressante serait donc de pouvoir les tester, ce qui dépend grandement de la disponibilité de données relatives aux déplacements, qui soient compatibles avec le référentiel fin des grilles régulières.

A côté de cette première perspective qui ne semble envisageable qu'à plus long terme, une perspective à court terme serait de tester l'apport des variables socio-économiques de l'INSEE qui semblent maintenant être diffusées de façon stable sur les grilles régulières. L'intégration de ces données pourrait en effet constituer un apport intéressant pour caractériser la morphologie socio-économique des aires urbaines, qui est elle aussi liée à la morphologie du bâti et aux formes d'interaction et pratiques de déplacement. De plus la mise à jour des données carroyées de l'INSEE, encore incertaines mais envisagée sur une base quinquennale¹³ offrira des perspectives intéressantes d'étude longitudinale, mais envisageable uniquement à long terme.

Perspectives méthodologiques

Du point de vue méthodologique, plusieurs perspectives sont envisageables pour améliorer le traitement des données sur grille régulière, mais aussi pour améliorer les représentations possibles des données. Concernant la méthode de classification par transfert d'apprentissage, il serait intéressant de pouvoir renforcer le processus de sélection des modèles de liens paramétriques entre les distributions statistiques des aires urbaines. Nous l'avons vu, les critères classiques tels que BIC et ICL ne donnent pas les résultats attendus et présentent une certaine instabilité. Une possibilité serait de

13. D'après les informations communiquées au Séminaire de Méthodologie Statistique du département des méthodes statistiques de l'INSEE du 20/01/2015.

s'inspirer des critères de sélection conçus pour des tâches de classification supervisée, tels que le critère BEC (*Bayesian Entropy Criterion*) de Bouchard et Celeux (2004) pour proposer un critère utilisant les approximations de log-vraisemblance disponibles.

D'autre part, il pourrait être intéressant de tester de nouveaux modèles de liens paramétriques pour modéliser de nouvelles formes de distorsion des covariables. Il serait judicieux de tester des modèles affines pour intégrer un degré de liberté supplémentaire, mais qui ne soit pas aussi grand que celui des modèles de la famille ($M5$) qui donnaient une trop grande flexibilité au classifieur. L'idée serait de proposer des modèles capables de tenir compte des particularités des classes, qui semblent influencer les distorsions dans l'application que nous avons considérée.

Enfin, une dernière idée pour améliorer l'algorithme de classification des tissus urbains serait de se tourner vers un mode de transfert d'apprentissage multi-tâche. Une idée serait par exemple d'utiliser les différentes sources d'apprentissage simultanément pour apprendre au mieux de nouvelles aires urbaines, qui correspondraient en fait à des situations intermédiaires entre ces prototypes. Une autre serait de pouvoir réaliser les classifieurs appris correctement au fur et à mesure pour stabiliser les ensembles d'apprentissage et améliorer l'apprentissage des tâches à venir.

En ce qui concerne la seconde phase de classification des données selon la nomenclature de morphotypes représentant des instances du modèle chorématique, les perspectives identifiées concernent les représentations et méthodes d'analyse associées. Une des idées proposées au tout début de cette thèse était de travailler sous des formats image et graphe. L'avantage de travailler à partir de graphes serait de pouvoir intégrer simultanément des informations relatives aux régions d'images qui constitueraient les noeuds des graphes, et des informations relatives aux interactions entre ces régions qui constitueraient les arcs des graphes. Cette représentation permettrait par exemple d'étudier des liens entre la morphologie et les pratiques de déplacement. Une fois les obstacles liés à la disponibilité de données de mobilité levés, à nos yeux la perspective la plus intéressante serait donc de pouvoir développer ce type de représentation. Cette représentation, en plus de permettre une valorisation inédite des liens entre configurations morphologiques et fonctionnelles soulève des questions méthodologiques particulièrement intéressante sur l'analyse et la comparaison de graphes aux noeuds et aux arcs valués .

Liste des figures

1	Structure de la thèse	12
1.1	Modèle de Von Thünen (base graphique tirée du Oxford Dictionary of Geography)	20
1.2	Modèle auréolaire de Burgess sur la ville de Chicago, Source : (Burgess, 1925, p. 55)	21
1.3	Modèle sectoriel de Hoyt, Source : Harris et Ullman (1945)	25
1.4	Modèle des noyaux multiples de Harris et Ullman, Source : Harris et Ullman (1945)	27
1.5	Modèle d'occupation spatiale issu de la géographie structurale, Source : Ritchot (1992)	29
1.6	Modèles d'organisation des lieux centraux de Christaller	37
1.7	Modèle d'organisation Fractalopolis (selon un tapis de Sierpinski multi-fractal)	42
2.1	Population résidente sur la grille à mailles de 200 mètres de côté de la France	59
2.2	Grilles de population sur la communauté d'agglomération du Grand Besançon : 2.2a Mailles d'1 km de côté, 2.2b Mailles de 200 mètres de côté	60
2.3	Modélisations géométriques des bâtiments de la BD Topo®	65
2.4	Variations de l'entropie selon les configurations de quartiers observées sur l'aire urbaine Lyonnaise	69
2.5	Répartition de la population sur la grille de population à maille de 200 mètres de côté pour le quartier de Parilly (Bron, 69)	72
2.6	Règles de décision pour la classification des tissus urbains	80
2.7	Profils de densités sur l'agglomération de Besançon : a) profil 47, b) profil 57 (Habillage : BD Topo®, IGN)	82
2.8	Dendrogramme de Besançon (coupé à la partition en 16 groupes)	84
2.9	Tissus urbains et secteurs d'EMD : Communauté d'agglomération du Grand Besançon	85
2.10	Tissus urbains et secteurs d'EMD : SCOT de l'Agglomération Tourangelle	86
2.11	Tissus urbains et secteurs d'EMD : SCOT Sud 54	87
3.1	Systèmes de voisinages 4-D et 8-D et cliques associées	108

3.2	Images simulées en utilisant un modèle de Potts à 6 classes : (a) Image à structure aléatoire (β_{rand}), (b) Image à groupes concentrés (β_{struc}) . . .	123
3.3	Densités de probabilité marginales des variables de population, surface bâtie et entropie pour des modèles de mélange appris sur les différentes sources d'apprentissage qui seront utilisées pour constituer la base de tissus urbains	127
3.4	Valeurs absolues des corrélations entre variables pour les quatre départements incluant les sources d'apprentissage	130
3.5	Importances globales des variables	131
3.6	Erreurs OOB de classifications sur les quatre jeux d'apprentissage	132
3.7	Images classées de Besançon selon différentes situations d'apprentissage depuis Tours	136
3.8	Images classées de Lyon selon différentes situations d'apprentissage depuis Strasbourg	138
3.9	Images classées de Tours selon différentes situations d'apprentissage depuis Lyon	139
3.10	Images classées de Besançon selon différentes situations d'apprentissage depuis Strasbourg	141
3.11	Comportement de BIC face à l'erreur de classification	143
3.12	Comportement de ICL face à l'erreur de classification	144
3.13	Cartographie de l'aire urbaine de Nantes : a) MOS, b) classification en tissus urbains	146
3.14	Cartographie de l'aire urbaine de Strasbourg : a) MOS, b) classification en tissus urbains	149
4.1	Table chorématique de Brunet (1986)	161
4.2	Chorèmes urbains	163
4.3	Représentation chorématique générique d'une forme urbaine métropolitaine (sans prise en compte de contraintes spatiales influençant les processus de hiérarchisation)	165
4.4	Contraintes spatiales générant des particularités d'organisation	167
4.5	Modèle chorématique de l'aire urbaine de Besançon	169
4.6	Modèle chorématique de l'aire urbaine de Lyon	170
4.7	Éléments structurants utilisés	173
4.8	Illustration des relations topologiques du modèle RCC-8	175
4.9	Post-traitement de simplification à l'échelle du pixel	181
4.10	Post-traitement de simplification à l'échelle de la région	182
4.11	Étapes de simplification de l'image classée des tissus urbains de Lorient	183
4.12	Identification des formes d'habitat individuel coalescente sur l'agglomération de Mulhouse	184
4.13	Règles de décision pour passer des tissus urbains aux morphotypes . . .	186

4.14	Organisation morphologique de l'aire urbaine de Lille (partie française) : a) Modèle chorématique, b) Classification en morphotypes	189
4.15	Organisation morphologique de l'aire urbaine de Toulouse : a) Modèle chorématique, b) Classification en morphotypes	190
5.1	Proportions de population par morphotypes résidentiels pour les 50 aires urbaines de l'échantillon	198
5.2	Proportions de la fragmentation des régions par morphotypes de chaque aire urbaine de l'échantillon	200
5.3	Matrices de relations spatiales dans l'aire urbaine de Lille	201
5.4	Matrices de relations spatiales dans l'aire urbaine de Toulouse	201
5.5	Evaluation de la forme de l'agglomération de Besançon	203
5.6	Ensemble des matrices traitées par AFM et notations	210
5.7	Premier plan factoriel de l'ACP sur les indices de forme	217
5.8	Premier plan factoriel de l'AFC sur la répartition de la population . . .	219
5.9	Premier plan factoriel de l'AFC fragmentation intra-urbaine	222
5.10	Premier plan factoriel de l'AFC relations intra-urbaines	224
5.11	Premier plan factoriel de l'AFMM	229
5.12	Arbre hiérarchique des aires urbaines (partition en 11 groupes)	231
5.13	Distance de Ward en fonction du nombre de groupes	232
5.14	Chorotypes de structure urbaine et organisation hiérarchique interclasse correspondant à la partie supérieure du dendrogramme	238

Liste des tables

1.1	Caractéristiques et rapports d'opposition pour un modèle morphologique	53
2.1	Définitions d'indicateurs de densité de contenant calculés dans un voisinage d'ordre 1	68
2.2	Hypothèses sur les valeurs relatives des indicateurs qui permettront de détecter différents tissus urbains	79
2.3	Seuils pour la classification des tissus urbains selon les agglomérations	81
3.1	Différentes approches de transfert d'apprentissage	103
3.2	Modèles de transfert dans le cadre markovien	117
3.3	Paramètres de distribution des covariables de l'image I_S	123
3.4	Erreurs de classification par apprentissage supervisé	124
3.5	Taux d'erreur de classification sur l'image <i>Cible</i> dans différentes conditions d'apprentissage	125
3.6	Distances aux vrais paramètres de l'image <i>Cible</i>	126
3.7	Bases de labels (nombre de pixels)	129
3.8	Valeurs moyennes de classes pour les quatres jeux d'apprentissages	133
3.9	Erreurs de classification selon les voisinages (4-D ou 8-D contraint) et l'estimation du champ voisin	134
3.10	Légende des images classées en tissus urbains	135
3.11	Erreurs de classification dans différentes situations d'apprentissage de Besançon depuis Tours	136
3.12	Matrices de confusion (en nombre de pixels) des modèles de Besançon appris par trNREM depuis Tours	137
3.13	Erreurs de classification dans différentes situations d'apprentissage de Lyon depuis Strasbourg	137
3.14	Matrices de confusion (en nombre de pixels) des modèles de Lyon appris par trNREM depuis Strasbourg	138
3.15	Erreurs de classification dans différentes situations d'apprentissage de Tours depuis Lyon	140
3.16	Matrices de confusion (en nombre de pixels) des modèles de Tours appris par trNREM depuis Lyon	140
3.17	Matrices de confusion (en nombre de pixels) des modèles de Besançon appris par trNREM depuis Strasbourg	141

3.18	Matrices de confusion (en nombre de pixels) des modèles de Besançon appris par trNREM depuis Strasbourg	142
3.19	Table de contingence entre les classes de tissus urbains et la BD MOS sur l'aire urbaine de Nantes	147
3.20	Table de contingence entre les classes de tissus urbains et la BD MOS sur l'aire urbaine de Strasbourg	148
3.21	Aires urbaines composant la base de données des tissus urbains	151
4.1	Relations topologiques définissables à partir de la primitive C	176
4.2	Relations topologiques considérées dans notre représentation	180
5.1	Labels des relations d'adjacence étudiées	202
5.2	Ensemble des variables composant l'espace de caractéristique des aires urbaines	205
5.3	Méthodologies comparées de l'AS et l'AFMTC	212
5.4	Décomposition de l'inertie par axe de l'ACP sur indices de forme	216
5.5	Décomposition de l'inertie par axe de l'AFC répartition de population	218
5.6	Décomposition de l'inertie par axe de l'AFC fragmentation intra-urbaine	221
5.7	Décomposition de l'inertie par axe de l'AFC relations intra-urbaines	223
5.8	Décomposition de l'inertie par axe de l'AFM	227
5.9	Contributions des groupes de variables à la création des axes de l'AFMM	228
5.10	Corrélations entre groupes de variables	228
5.11	Profils des groupes de la CAH : moyenne et écart-type des composantes globales	234

Notations

Sigles et acronymes

ACM	Analyse des correspondances multiples
ACP	Analyse en Composantes Principales
AFC	Analyse Factorielle des Correspondances
AFM	Analyse Factorielle Multiple
AFMM	Analyse Factorielle Multiple de données Mixtes
AFMTC	Analyse Factorielle Multiple de Tableaux de Contingence
AS	Analyse Simultanée
CAH	Classification Ascendante Hiérarchique
CART	<i>Classification And Regression Tree</i>
CBD	<i>Central Business District</i>
CERTU	Centre d'études sur les réseaux, les transports, l'urbanisme et les constructions publiques
CLC	Corine Land Cover
EEA	Agence Européenne pour l'Environnement
EMD	Enquête Ménage-Déplacement
EPCI	Établissement Public de Coopération Inter-communale
GMM	<i>Gaussian Mixture Model</i>
HMM	<i>Hidden Markov Model</i>
IGN	Institut national de l'information géographique et forestière
INSEE	Institut National de la Statistique et des Études Économiques
LUZ	<i>Large Urban Zone</i>
MAP	<i>Maximum a posteriori</i>
MOS	Mode d'Occupation des Sols
OCDE	Organisation de Coopération et de Développement Économiques
PIAO	Photo-Interprétation Assistée par Ordinateur
RFL	Revenus Fiscaux Localisés
SCoT	Schéma de Cohérence Territoriale
UMZ	<i>Urban Morphological Zone</i>

Variables et domaines

i.d.d.	(variables) indépendantes et identiquement distribuées
\mathbf{x}	Jeu de données
n	nombre d'individus dans \mathbf{x}
p	dimension du vecteur $x \in \mathbf{x}$
\mathbf{x}_S	Jeu de données <i>Source</i>
\mathbf{x}_T	Jeu de données <i>Cible</i>
\mathbf{y}	Vecteur d'étiquettes
\mathbf{y}_S	Vecteur d'étiquettes <i>Source</i>
\mathbf{y}_T	Vecteur d'étiquettes <i>Cible</i>
\mathcal{A}	Ensemble d'entraînement (apprentissage supervisé)
$\mathcal{D}_S = \{\mathbf{x}_S, P_S(\mathbf{x})\}$	Domaine <i>Source</i>
$\mathcal{D}_T = \{\mathbf{x}_T, P_T(\mathbf{x})\}$	Domaine <i>Cible</i>
$f_S(\cdot)$	Fonction prédictive générique sur un domaine <i>Source</i>
$f_T(\cdot)$	Fonction prédictive générique sur un domaine <i>Cible</i>
$\mathcal{T}_S = \{\mathbf{y}_S, f_S(\cdot)\}$	Tâche <i>Source</i>
$\mathcal{T}_T = \{\mathbf{y}_T, f_T(\cdot)\}$	Tâche <i>Cible</i>

Fonctions et probabilités

$D(A, B)$	Critère générique d'agrégation de groupes d'individus
$d(x_1, x_2)$	Distance générique entre deux individus
$\lambda(\mathbf{x}; \mathcal{A})$	Arbre de classification
$P(\cdot)$	Notation générique d'une fonction de probabilité
$f(\cdot, \theta)$	Notation générique d'une fonction de densité θ
$\mathcal{N}(\mu, \Sigma)$	Loi normale de moyenne μ et de matrice de variance-covariance Σ
$E(\mathbf{x})$	Espérance de \mathbf{x}
$L(\mathbf{x}, \psi)$	Fonction de vraisemblance associée au paramètre ψ par rapport au jeu de données \mathbf{x}
$\mathcal{L}(\mathbf{x}, \psi)$	Fonction de log-vraisemblance associée au paramètre ψ par rapport au jeu de données \mathbf{x}
$\mathcal{PL}(\mathbf{x}, \psi)$	Fonction de pseudo-vraisemblance de Besag (1975) associée au paramètre ψ par rapport au jeu de données \mathbf{x}

Algèbre linéaire

X^t	Transposée de la matrice X
X^{-1}	Inverse de la matrice X
$ X $	Déterminant de la matrice X
\bar{x}	Vecteur moyen

Champs de Markov

I	Image composée de n pixels
$\mathcal{N}(I)$	Système de voisinage associé à l'image I
$N(i)$	Ensemble des pixels voisins du pixel i
$i \sim j$	Paire de pixels voisins
$H(\cdot)$	Notation générique d'une fonction d'énergie
$V_c(\cdot)$	Notation générique d'une fonction potentielle associée à des cliques c
$W(\mathbf{y})$	Constante de normalisation du champ de Markov \mathbf{y}
\mathbf{y}^{mf}	Champ voisin
$P_{\mathbf{y}^{mf}}(\mathbf{y})$	Fonction de probabilité du champ de Markov \mathbf{y} sous l'approximation en champ moyen

Connecteurs logiques

\neg	Négation logique (Non)
\wedge	Conjonction logique (Et)
\vee	Disjonction logique (Ou)
\rightarrow	Implication logique (Si ... Alors ...)
\equiv	Équivalence logique (Si et seulement si)

Analyse factorielle

$CL_k(i)$	Notation générique de la k -ème composante principale projetant l'individu i par analyse factorielle
$CC_k(j)$	Notation générique de la k -ème composante principale projetant la variable j par analyse factorielle
$Ctr_k[g]$	Contribution du groupe de variables g à la création de la k -ème composante principale
$Ctr_k(j)$	Contribution de la variable j à la création de la k -ème composante principale
f_{ij}	Élément à la i -ème ligne et j -ème colonne de la matrice de fréquence F
$f_{i.}^g$	Vecteur de fréquence marginale par ligne pour le g -ème groupe de F
$f_{i..}$	Vecteur de fréquence marginale calculée globalement sur les lignes de F
$f_{.j}^g$	Vecteur de fréquence marginale pour la j -ème colonne du g -ème groupe de F
i^g	Profil-ligne du i -ème individu étudié uniquement à partir des variables du groupe g
i^c	Compromis du profil-ligne du i -ème individu étudié à partir de tous les groupes de variables

Bibliographie

- P. Abrantes, C. Soulard, F. Jarrige, et L. Laurens. Dynamiques urbaines et mutations des espaces agricoles en Languedoc-Roussillon (France). *Cybergeo : European Journal of Geography*, 485, 2010.
- M. Aiello, I. Pratt-Hartmann, et J. V. Benthem. *Handbook of Spatial Logics*. Springer, 2007, 1058 p.
- R. Allain. *Morphologie urbaine. Géographie, aménagement et architecture de la ville*. Armand Colin, Collection U, Paris, 2004, 256 p.
- J. F. Allen. Maintining knowledge about temporal intervals. *Communications of the ACM*, 26(11) :832–843, 1983.
- W. Alonso. *Location and land use. Toward a general theory of land rent*. Harvard University Press, 1964, 201 p.
- W. Alonso. Urban zero population growth. *Daedalus*, 102(4) :191–206, 1973.
- A. Anas et I. Kim. Income distribution and the residential density gradient. *Journal of Urban Economics*, 31(2) :164–180, 1992.
- A. Anas, R. Arnott, et K. Small. Urban spatial structure. *Journal of Economic Literature*, 36(3) :1426–1464, 1998.
- A. Annoni, editor. *Proceedings of the 1st Workshop on European Reference Grids*, Ispra, 2003. JRC-Institute for Environment and Sustainability.
- F. Ascher. *Les nouveaux principes de l'urbanisme*. Editions de l'Aube, 2001, 104 p.
- ATU (Agence d'Urbanisme de l'agglomération de Tours). Questions de mobilité n° 4 : Comment et pourquoi se déplacent les habitants?, 2010.
- D. Badariotti. Des fractales pour l'urbanisme? quelques pistes de réflexion á partir de l'exemple de strasbourg-kehl. *Cahiers de géographie du Québec*, 49(137) :133–156, 2005.

- A. S. Bailly. Les théories de l'organisation de l'espace urbain. *Espace géographique*, 2(2) : 81–93, 1973.
- P. Balbiani et P. Muller. Le raisonnement spatial. In *Le temps, l'espace et l'évolutif en sciences du traitement de l'information*, pages 33–53. CEPAD, septembre 2000.
- D. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2) :111–122, 1981.
- D. Banister. *Transport, formes urbaines et croissance économique*, Chapitre Villes, forme et étalement urbain : Une perspective européenne, pages 121–147. Editions OCDE, 2007.
- D. Batten. Network cities : Creative urban agglomerations for the 21st century. *Urban Studies*, 32(2) :313–327, 1995.
- M. Batty, E. Besussi, K. Maat, et J. Harts. Representing multifunctional cities : Density and diversity in space and time. *Built Environment*, 30(4) :324–337, 2004.
- M. Batty et K. S. Kim. Form follows function : Reformulating urban population density functions. *Urban Studies*, 29(7) :1043–1070, 1992.
- M. Batty et P. Longley. The fractal simulation of urban structure. *Environment and Planning A*, 18(9) :1143–1179, 1986.
- G. Baudelle et E. Ducom. L'organisation de l'espace urbain par la distance au centre : des modèles contradictoires ? *Atala*, 12 :85–101, 2009.
- C. Baumont et J.-M. Huriot. La ville abstraite. une analyse des représentations théoriques de la ville. Document de travail 9501, LATEC, 1995.
- J.-J. Bavoux, F. Beaucire, L. Chapelon, et P. Zembri. *Géographie des transports*. Armand Colin, Collection U, Paris, 2005, 232 p.
- J. Beaujeu-Garnier. *Géographie urbaine*. Armand Colin, Collection U, 1995, 349 p.
- M. Bécue-Bertaut et J. Pagès. A principal axes method for comparing contingency tables : {MFACT}. *Computational Statistics & Data Analysis*, 45(3) :481 – 503, 2004.
- M. Bécue-Bertaut et J. Pagès. Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data. *Computational Statistics & Data Analysis*, 52(6) : 3255 – 3268, 2008.
- R. Bellman. *Adaptive Control Processes : A Guided Tour*. Princeton University Press, 1961, 255 p.
- F. Beninel, C. Biernacki, C. Bouveyron, J. Jacques, et A. Lourme. *Knowledge Transfer : Practices, Types and Challenges*, Chapitre Parametric link models for knowledge transfer in statistical learning. Nova Publishers, 2012.
- J.-P. Benzécri. *L'analyse des données : L'analyse des correspondances*. Dunod, 1973, 619 p.

- B. Berry. Cities as systems within systems of cities. *Papers of the Regional Science Association*, 13(1) :146–163, 1964.
- A. Bertaud. The spatial organization of cities : Deliberate outcome or unforeseen consequence?, 2004.
- J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2) :192–236, 1974.
- J. Besag. Statistical analysis of non-lattice data. *The statistician*, 24(3) :179–195, 1975.
- J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, 48(3) :259–302, 1986.
- P. Bessy-Pietri. Les formes récentes de la croissance urbaine. *Économie et statistique*, 6(336) : 35–52, 2000.
- C. Biernacki, G. Celeux, et G. Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7) :719–725, 2000.
- C. Biernacki, F. Beninel, et V. Bretagnolle. A generalized discriminant rule when training population and test population differ on their descriptive parameters. *Biometrics*, 58(2) : 387–397, 2002.
- I. Bloch. Fuzzy relative position between objects in image processing : A morphological approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7) :657–664, 1999.
- I. Bloch. *Applying Soft Computing in Defining Spatial Relations*, Chapitre Mathematical Morphology and Spatial Relationships : Quantitative, Semi-Quantitative and Symbolic Settings, pages 63–98. Physica Verlag Heidelberg, 2002.
- O. Bonin et B. Poupeau. "raisonnement spatial sur la position et l'orientation de solides dans l'espace. In *atelier RTE'06*, 2006.
- G. Borruzo. Network density and the delimitation of urban areas. *Transactions in GIS*, 7 (2) :177–191, 2003.
- M. Borzacchiello, P. Nijkamp, et E. Koomen. Accessibility and urban development : a grid-based comparative statistical analysis of dutch cities. *Environment and Planning B : Planning and Design*, 37(1) :148–169, 2010.
- G. Bouchard et G. Celeux. Model selection in supervised classification. Research report, INRIA, 2004.
- L. Bourdeau-Lepage, J.-M. Huriot, et J. Perreur. À la recherche de la centralité perdue. *Revue d'Économie Régionale et Urbaine*, 3 :549–572, 2009.

- C. Bouveyron et J. Jacques. Adaptive linear models for regression : Improving prediction when population has changed. *Pattern Recognition Letters*, 31(14) :2237 – 2247, 2010.
- S. Boyd et L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004, 716 p.
- J. S. Brainard, A. P. Jones, I. J. Bateman, A. A. Lovett, et P. J. Fallon. Modelling environmental equity : access to air quality in birmingham. *Environment and Planning A*, 34(4) : 695–716, 2002.
- F. Braudel. *Civilisation matérielle, économie et capitalisme, XVe-XVIIIe siècle*, Volume 3 - Le temps du monde. Armand Colin, 1979, 607 p.
- L. Breiman. Bagging predictors. *Machine Learning*, 24(2) :123–140, 1996.
- L. Breiman. Random forests. *Machine Learning*, 45(1) :5–32, 2001.
- L. Breiman, J. H. Friedman, R. A. Olshen, et C. J. Stone. *Classification and Regression Trees*. Statistics/Probability Series. Wadsworth Publishing Company, 1984.
- A. Bretagnolle, D. Pumain, et C. Vacchiani-Marcuzzo. The organization of urban systems. In *Complexity Perspectives in Innovation and Social Change*, pages 197–220. Springer Netherlands, 2009.
- G. H. Brundtland. Notre avenir à tous. Rapport, Commission mondiale sur l’environnement et le développement, 1987.
- R. Brunet. La carte-modèle et les chorèmes. *Mappemonde*, 86(4) :2–6, 1986.
- R. Brunet. La composition des modèles dans l’analyse spatiale. *L’Espace Géographique*, 9 (4) :253–265, 1980.
- R. Brunet. *La carte, mode d’emploi*. Fayard, 1987, 270 p.
- R. Brunet et H. Théry. *Les mots de la Géographie - Dictionnaire critique*. Reclus, Documentation Française, 1992.
- C. Brutel et D. Levy. Le nouveau zonage en aires urbaines de 2010. *INSEE Première*, 1374 : 4 p., 2011.
- C. Brutel et D. Levy. Le nouveau zonage en bassins de vie de 2012. *INSEE Première*, 1425 : 4 p., 2012.
- M. Burger et E. Meijers. Form follows function ? linking morphological and functional polycentricity. *Urban Studies*, 49(5) :1127–1149, 2012.
- E. Burgess. The growth of the city : An introduction to a research project. In *The City*. University of Chicago Press, 1925.
- CAGB (Communauté d’Agglomération du Grand Besançon). Enquête ménages déplacements. Résultats 2004-2005, 2006.

- R. Capello. The city network paradigm : Measuring urban network externalities. *Urban Studies*, 37(11) :1925–1945, 2000.
- R. Caruana. Multitask learning. In *Learning to Learn*, pages 95–133. Springer US, 1998.
- F. Catté, P.-L. Lions, J.-M. Morel, et T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 29(1) :182–193, 1992.
- CE (Commission Européenne). Livre vert sur l’environnement urbain. Rapprt, Commission des Communautés euroéennes, 1990.
- G. Celeux, D. Chauveau, et J. Diebolt. On stochastic versions of the EM algorithm. Research report, INRIA, 1995.
- G. Celeux, F. Forbes, et N. Peyrard. EM procedures using mean field-like approximations for markov model-based image segmentation. *Pattern Recognition*, 36(1) :131–144, 2003.
- G. Celeux, F. Forbes, et N. Peyrard. Modèle de potts avec champ externe et algorithme de type EM pour la segmentation d’image. In *RFIA*, Toulouse, 2004.
- R. Cervero et K. Kockelman. Travel demand and the 3Ds : Density, diversity, and design. *Transportation Research Part D : Transport and Environment*, 2(3) :199 – 219, 1997.
- R. Cervero et K.-L. Wu. Polycentrism, commuting, and residential location in the san francisco bay area. *Environment and Planning A*, 29(5) :865–886, 1997.
- B. Chalmond. An iterative gibbsian technique for reconstruction of m-ary images. *Pattern Recognition*, 22(6) :747–761, 1989.
- D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, 1987, 274 p.
- E. Charmes. La densification en débat, Effet de mode ou solution durable ? *Etudes foncières*, 145 :19–38, 2010.
- Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8) :790–799, 1995.
- W. Christaller. *Die Zentralen Orte in Süddeutschland*. Iena, Fischer, 1933, 331 p.
- J. R. Cladera, C. R. M. Duarte, et M. Moix. Urban structure and polycentrism : Towards a redefinition of the sub-centre concept. *Urban Studies*, 46(13) :2841–2868, 2009.
- C. Clark. Urban population densities. *Journal of the Royal Statistical Society. Series A*, 114 (4) :490–496, 1951.
- B. J. Clarke. A calculus of individuals based on ‘connection’. *Notre Dame Journal of Formal Logic*, 22(3) :204–218, 1981.

- E. Clementini et P. D. Felice. Approximate topological relations. *International Journal of Approximate Reasoning*, 16(2) :173–204, 1997.
- E. Clementini, J. Sharma, et M. J. Egenhofer. Modeling topological spatial relations : Strategies for query processing. *Computers and Graphics*, 18(6) :815–822, 1994.
- J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1) :37–46, 1960.
- A. G. Cohn et J. Renz. *Handbook of Knowledge Representation*, Chapitre Qualitative Spatial Representation and Reasoning, pages 551–596. Elsevier, 2008.
- A. G. Cohn, B. Bennett, J. Gooday, et N. M. Gotts. Qualitative spatial representation and reasoning with the region connection calculus. *GeoInformatica*, 1 :275–316, 1997.
- O. Colliot. *Représentation, évaluation et utilisation de relations spatiales pour l'interprétation d'images, Application à la reconnaissance de structures anatomiques en image médicale*. Thèse de doctorat, Ecole nationale supérieure des télécommunications - Département TSI, 2003.
- E. Côme. *Apprentissage de modèles génératifs pour le diagnostic de systèmes complexes avec labellisation douce et contraintes spatiales*. Thèse de doctorat, Université de Technologie de Compiègne, 2009.
- R. Crane. The influence of urban form on travel : An interpretive review. *Journal of Planning Literature*, 15(1) :3–23, 2000.
- J. W. Creswell et V. L. P. Clark. *Designing and Conducting Mixed Methods Research*. Sage Publications, Inc, 2nd edition, 2010, 488 p.
- M. Cristani, A. G. Cohn, et B. Bennett. Spatial locations via morpho-mereology. In *Principles of Knowledge Representation and Reasoning, KR2000*, 2000.
- E. Cuniberti, C. Frandon, et E. Giraud. Les bases de données géographiques d'occupation du sol, volet tache urbaine. Rapport, CERTU, 2005.
- H. Daumé III et D. Marcu. Domain adaptation for statistical classifiers. *Journal Of Artificial Intelligence Research*, 26 :101–126, 2006.
- M.-L. De Keersmaecker, P. Frankhauser, et I. Thomas. Dimensions fractales et réalités péri-urbaines. l'exemple du sud de bruxelles. *L'espace Géographique*, 33(3) :219–240, 2004.
- P. De Roo. Villes moyennes, villes d'intermédiation, pour une diversité des modèles de développement urbain. Livre blanc, DATAR, 2005.
- M. J. de Smith, M. F. Goodchild, et P. A. Longley. *Geospatial Analysis. A Comprehensive Guide to Principles, Techniques and Software Tools, 5th Edition*. The Winchelsea Press, 2014, 748 p.

- A. P. Dempster, N. M. Laird, et D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 39(1) :1–38, 1977.
- G. Desmarais et G. Ritchot. *La géographie structurale*. L'harmattan, 2000.
- DIACT. *Les villes moyennes françaises. Enjeux et perspectives*. Collection Travaux. La documentation française, 2007, 64 p.
- R. Diaz-Uriarte et S. A. de Andres. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7(3), 2006.
- L. Dijkstra et H. Poelman. Cities in europe - the new OECD-EC definition. *Regional Focus (European Commission)*, 1 :16 p., 2012.
- G. Djament-Tran et C. Grataloup. E pluribus urbibus una : Modéliser les trajectoires de villes. *Mappemonde*, 100(4) :1–13, 2010.
- C. Do et A. Ng. Transfer learning for text classification. In *Proceeding of Advances in Neural Information Processing Systems, NIPS*, Volume 18, December 5-8 2005.
- R. O. Duda, P. E. hart, et D. G. Stork. *Pattern Classification*. Wiley, 2000, 680 p.
- M. J. Egenhofer et R. D. Franzosa. Point-set topological spatial relations. *International Journal of Geographical Information Systems*, 5(2) :161–174, 1991.
- L. Eriksson, E. Johansson, N. Kettaneh-Wold, et S. Wold. *Multi- and Megavariate Data Analysis : Principles and Applications*. Umetrics Academy, 2001.
- B. Escoffier et J. Pagès. *Analyses factorielles simples et multiples*. Dunod, 2008, 328 p.
- Eurostat. *Eurostat regional yearbook 2010*, Chapitre A revised urban-rural typology, pages 241–255. Publications Office of the European Union, 2010.
- R. Ewing et R. Cervero. Travel and the built environment. *Journal of the American Planning Association*, 76(3) :265–294, 2010.
- I. K. Fodor. A survey of dimension reduction techniques. Technical Report UCRL-ID-148494, Lawrence Livermore National Laboratory, 2002.
- F. Forbes et N. Peyrard. Hidden markov random field model selection criteria based on mean field-like approximations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9) :1089–1101, 2003.
- V. Fouchier. *Les densités urbaines et le développement durable, le cas de l'Ile de France et des villes nouvelles*. Secrétariat Général du Groupe central des Villes Nouvelles, 1997.
- P. Frankhauser. *La fractalité des structures urbaines*. Anthropos, 1994, 291 p.
- P. Frankhauser. The fractalopolis model - a sustainable approach for a central place system. Research report, ThéMa, 2012.

- P. Frankhauser, H. Houot, C. Tannier, et G. Vuidel. Vers des d'éplacements périurbains plus durables : proposition de modèles fractals opérationnels d'urbanisation. Rapport final PREDIT 3, ThéMa, 2007.
- M. Fujita et J.-F. Thisse. *Économie des villes et de la localisation*. De Boeck Université, 2003, 559 p.
- M. Fujita, P. Krugman, et A. J. Venables. *The Spatial Economy : Cities, Regions, and International Trade*. The MIT Press, 1999.
- J. Garreau. *Edge City : Life on the New Frontier*. Anchor Books, 1991, 548 p.
- F. Gaschet. The new intra-urban dynamics : Suburbanisation and functional specialisation in french cities. *Papers in Regional Science*, 81(1) :63–81, 2002.
- F. Gaschet et C. Lacour. Métropolisation, centre et centralité. *Revue d'économie Régionale & Urbaine*, 1 :49–72, 2002.
- M. Gascon. *L'enquête ménages déplacements « Standard Certu » : Guide méthodologique*. Editions du CERTU, 2008.
- B. Gauthiez. *Espace urbain, vocabulaire et morphologie*. Editions du Patrimoine, 2003, 494p.
- S. Geman et D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6 (6) :721–741, 1984.
- R. Genuer, J.-M. Poggi, et C. Tuleau-Malot. Variable selection using random forests. *Pattern Recognition Letters*, 31(14) :2225–2236, 2010.
- J. Gill. *Generalized Linear Models : A Unified Approach*. SAGE Publications, 2001, 101 p.
- G. Giuliano et C. Redfearn. Not all sprawl : Evolution of employment concentrations in los angeles, 1980-2000. Technical report, School of Policy, Planning and Development, USC, 2005.
- G. Giuliano et K. A. Small. Subcenters in the los angeles region. *Regional Science and Urban Economics*, 21(2) :163–182, 1991.
- G. Giuliano, C. Redfearn, A. Agarwal, et S. He. Network accessibility and employment centres. *Urban Studies*, 49(1) :77–95, 2012.
- P. Gordon et H. W. Richardson. Beyond polycentricity : The dispersed metropolis, los angeles, 1970-1990. *Journal of the American Planning Association*, 62(3) :289–295, 1996.
- P. Gordon, A. Kumar, et H. W. Richardson. The influence of metropolitan spatial structure on commuting time. *Journal of urban economics*, 26(2) :138–151, 1989.
- P. Gordon, H. W. Richardson, et M.-J. Jun. The commuting paradox evidence from the top twenty. *Journal of the American Planning Association*, 57(4) :416–420, 1991.

- J. Gottmann. *Megalopolis. The urbanized North-eastern seaboard of the United States*. The 20th Century Fund, 1961.
- C. Grataloup et D. Eckert. Chrono-chorématique urbaine. *Mappemonde*, 100(4) :–, 2010.
- B. Grosjean. *Projets et stratégies urbaines : regards comparatifs*, Chapitre Construction et maîtrise des formes urbaines : lecture comparée des SCoT de Lille, Nantes, Bordeaux et Montpellier, pages 238–257. Editions Parenthèses, 2009.
- J. L. Gross et J. Yellen. *Handbook of Graph Theory*. CRC Press, 2004, 1192 p.
- M. Guérois et F. Paulus. Commune centre, agglomération, aire urbaine : quelle pertinence pour l'étude des villes ? *Cybergeo : European Journal of Geography*, 212 :1–20, 2002.
- M. Guérois et D. Pumain. Built-up encroachment and the urban field : a comparison of forty european cities. *Environment and Planning A*, 40 :2186–2203, 2008.
- P. H. C. Guerra, A. Veloso, W. Meira, et V. Almeida. From bias to opinion : a transfer-learning approach to real-time sentiment analysis. In *Proceedings of the 16th ACM SIGKDD conference*, pages 1159–1168, 2010.
- I. Guyon et A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3 :1157–1182, 2003.
- P. Haggett. *L'analyse spatiale en géographie humaine*. Armand Colin, Collection U, 1973.
- J. Han et M. Kamber. *Data Mining : Concept and Techniques Second Edition*. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann Publishers, 2006.
- C. D. Harris et E. L. Ullman. The nature of cities. *Annals of the American Academy of Political and Social Science*, 242 :7–17, 1945.
- T. Hastie, R. Tibshirani, et J. Friedman. *The Elements of Statistical Learning*. Springer. Springer New York Inc., 2nd edition, 2009, 745.
- W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1) :97–109, 1970.
- E. Heikkila, P. Gordon, J. I. Kim, R. B. P. harry W. Richardson, et D. Dale-Johnson. What happened to the cbd-distance gradient ? : land values in a policentric city. *Environment and Planning A*, 21(2) :221–232, 1989.
- M. Herold, J. Scepan, et K. C. Clarke. The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. *Environment and Planning A*, 34 : 1443–1458, 2002.
- P. M. Hess, A. V. Moudon, et M. Logsdon. Measuring land use patterns for transportation research. *Transportation Research Record*, 1780(1) :17–24, 2001.

- M. Hilal, A. Barczak, F.-P. Tourneux, Y. Schaeffer, M. Houdart, et D. Cremer-Schulte. Typologie des campagnes françaises et des espaces à enjeux spécifiques (littoral, montagne et dom). Travaux en Ligne 12, Datar, 2012.
- B. Hillier et L. Vaughan. The city as one thing. *Progress in Planning*, 67(3) :205–230, 2007.
- H. Hotelling. Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.*, 24(6) :417–441, 1933.
- H. Hoyt. *The structure and growth of residential neighborhoods in American cities*. Washington, Government printing office, 1939, 178 p.
- J. Huang, X. Lu, et J. Sellers. A global comparative analysis of urban form : Applying spatial metrics and remote sensing. *Landscape and Urban Planning*, 82 :184–197, 2007.
- J.-P. Hubert et F. Delisle. L’allongement des déplacements quotidiens contribue à l’émergence d’espaces urbains multipolaires, tandis que la mobilité baisse au centre des grandes agglomérations. In *La mobilité des Français, panorama issu de l’enquête nationale transports et déplacements 2008*. La Revue du CGDD, 2010.
- W. Hucy. Modéliser la nature en ville : expérimentation. *Nature Sciences Sociétés*, 18 : 133–146, 2010.
- J.-M. Huriot et J. Perreur. *L’intégration régionale des espaces*, Chapitre Centre et périphéries : fondements et critères, pages 63–84. Economica, 1997.
- J. Inglada et J. Michel. Qualitative spatial reasoning for high-resolution remote sensing image analysis. *Geoscience and Remote Sensing, IEEE Transactions*, 47(2) :599–612, 2009.
- W. Isard. *Location and Space-Economy*. Cambridge, The MIT Press, 1956, 350 p.
- E. Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1) :253–258, 1925.
- Y. R. Jabareen. Sustainable urban forms. their typologies, models, and concepts. *Journal of Planning Education and Research*, 26(1) :38–52, 2006.
- M. Jenks, E. Burton, et K. Williams. *The Compact City : A Sustainable Urban Form ?* Routledge, 1996.
- A. Katsaras. Fuzzy proximity spaces. *Journal of Mathematical Analysis and Applications*, 68 (1) :100–110, 1979.
- E. Koomen, P. Rietveld, et F. Bacao. The third dimension in urban geography : the urban-volume approach. *Environment and Planning B*, 36(6) :1008–1025, 2009.
- B. Kuipers. Modeling spatial knowledge. *Cognitive Science*, 2 :129–153, 1978.
- Y. Lacoste. Les géographes, la science et l’illusion. *Hérodote*, 76 :3–21, 1995.

- G. Lajoie. *Le carroyage des informations urbaines : une nouvelle forme de banque de données sur l'environnement du Grand Rouen*. Publications de l'université de Rouen, 1992, 238 p.
- R. E. Lang et J. LeFurgy. Edgeless cities : Examining the noncentered metropolis. *Housing Policy Debate*, 14(3) :427–460, 2003.
- S. Lardon et M. Capitaine. Chorèmes et graphes. Production et transformation de représentations spatiales en agronomie. *Revue d'anthropologie des connaissances*, 2(2) :195–217, 2008.
- F. Le Néchet. Consommation d'énergie et mobilité quotidienne selon la configuration des densités dans 34 villes européennes. *Cybergeo : European Journal of Geography*, 529 :27, May 2011.
- A. Lebreton et G. Mougel. La gentrification comme articulation entre forme urbaine et globalisation : approche comparative Londres/Berlin. *Espaces et sociétés*, 1(132-133) : 57–73, 2008.
- J. A. Lee et M. Verleysen. *Nonlinear Dimensionality Reduction*. Information Science and Statistics. Springer, 2007.
- S.-Y. Lee et F.-J. Hsu. Spatial reasoning and similarity retrieval of images using 2D C-string knowledge representation. *Pattern Recognition*, 25(3) :305–318, 1992.
- S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer London, 2009, 357 p.
- P. Livet, J.-P. Müller, D. Phan, et L. Sanders. Ontology, a mediator for agent-based modeling in social science. In *EPOS 08 3rd Edition of Epistemological Perspectives on Simulation A Cross Disciplinary Workshop*, 2008.
- P. Longley et V. Mesev. On the measurement and generalisation of urban form. *Environment and Planning A*, 32 :473–488, 2000.
- P. Longley et V. Mesev. Measurement of density gradients and space-filling in urban systems. *Papers in Regional Science*, 81(1) :1–28, 2002.
- A. Lösch. *The Economics of Location*. Yale University Press, 1954, 520 p.
- T. Louail. *Comparer les morphogenèses urbaines en Europe et aux Etats-Unis par la simulation à base d'agents*. Thèse de doctorat, Université d'Evry, 2010.
- T. Louail, M. Lenormand, O. G. Cantu Ros, M. Picornell, R. Herranz, E. Frias-Martinez, J. J. Ramasco, et M. Barthelemy. From mobile phone data to the spatial structure of cities. *Scientific Report*, 4(5276) :12 p., 2014.
- B. B. Mandelbrot. *The Fractal Geometry of Nature*. W.H.Freeman & Co Ltd, 1982, 460 p.
- S. Marshall. Urban pattern specification. Wp4 deliverable report, Bartlett School of Planning, University College London, 2009.

- D. Martin. Gridded population data for the UK redistribution models and applications. In *New Techniques and Technologies for Statistics Conference*, Brussels, Belgium, 2009.
- J. F. McDonald. The identification of urban employment subcenters. *Journal of Urban Economics*, 21(2) :242–258, 1987.
- J. F. McDonald et D. P. McMillen. Employment subcenters and land values in a polycentric urban area : the case of chicago. *Environment and Planning A*, 22(12) :1561–1574, 1990.
- K. McGarigal, S. Cushman, M. Neel, et E. Ene. FRAGSTATS : Spatial pattern analysis program for categorical maps. Technical report, University of Massachusetts, Amherst, 2002.
- G. McLachlan et D. Peel. *Finite Mixture Models*. Wiley, 2000, 456 p.
- E. Meijers. From central place to network model : Theory and evidence of a paradigm change. *Tijdschrift voor economische en sociale geografie*, 98(2) :245–259, 2007.
- E. Meijers. Summing small cities does not make a large city : Polycentric urban regions and the provision of cultural, leisure and sports amenities. *Urban Studies*, 45(11) :2323–2342, 2008.
- S. Mercier, R. Duthion, I. Boulet, et M. Buret. Les signes de la diversité toulousaine. *6 pages de l'INSEE*, 115 :1–6, 2008.
- C. Mering. Présentation de quelques méthodes de la morphologie mathématique permettant de caractériser une structure sur une image binaire. In *Journées Télédétection, Images satellite et milieux terrestres en régions arides et tropicales*, pages 193–211, 1990.
- P. Merlin et F. Choay. *Dictionnaire de l'urbanisme et de l'aménagement*. Presses Universitaires de France, 1988, 724 p.
- F. Meyer et S. Beucher. Morphological segmentation. *Journal of Visual Communication and Image Representation*, 1(1) :21–46, 1990.
- D. Mignot, A. Aguilera, et D. Bloy. Permanence des formes de la métropolisation et de l'étalement urbain. Rapport final, ADEME, 2004.
- D. Mignot, A. Aguilera, D. Bloy, D. Caubel, J.-L. Madre, L. Proulhac, et F. Vanco. Formes urbaines, mobilités et ségrégation :une comparaison lille-lyon-marseille. Rapport final, CPER-GRRT Socio-économie, 2007.
- H. J. Miller et J. Han. *Geographic Data Mining and Knowledge Discovery, 2nd Edition*. CRC Press, 2009, 486 p.
- E. S. Mills. An aggregative model of resource allocation in a metropolitan area. *The American Economic Review*, 57(2) :197–210, 1967.
- F. Moriconi-Ebrard. *GEOPOLIS - Pour comparer les villes du monde*. Economica, Collection Villes, 1994.

- R. F. Muth. *Cities and housing : the spatial pattern of urban residential land use*. Chicago, University of Chicago Press, 1969, 355 p.
- P. Naess. *Urban Structure Matters : Residential location, car dependence and travel behavior*. Routledge, 2006.
- P. Newman et J. Kenworthy. *Cities and automobile dependence : An international sourcebook*. Gower Publishing, 1989, 388 p.
- OCDE. *Panorama des régions de l'OCDE 2013*, Chapitre Définir les régions et les zones urbaines fonctionnelles, pages 1–14. Editions OCDE, 2013.
- S. Openshaw. *The modifiable areal unit problem*. Number 38 in Concepts And Techniques in Modern Geography (CATMOG). Geo Books, 1984, 41 p.
- S. J. Pan et Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10) :1345–1359, 2010.
- S. J. Pan, J. T. Kwok, et Q. Yang. Transfer learning via dimensionality reduction. In *23rd AAAI Conference on Artificial Intelligence*, 2008.
- J. Parr. A population-density approach to regional spatial structure. *Urban Studies*, 22 : 289–303, 1985.
- J. Parr. The polycentric urban region : A closer inspection. *Regional Studies*, 38(3) :231–240, 2004.
- K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6) :559–572, 1901.
- P. Perona et J. Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7) :629–639, Jul 1990.
- N. Peyrard. *Approximation de type champ moyen des modèles de champ de Markov pour la segmentation de données spatiales*. Thèse de doctorat, Université Joseph Fournier, 2001.
- W. Pieczynski. Champs de markov cachés et estimation conditionnelle itérative. *Traitement du Signal*, 11(2) :141–153, 1994.
- W. Pieczynski. Modèles de markov en traitements d'images. *Traitement du Signal*, 20(3) : 255–278, 2003.
- V. Piveteau et S. Lardon. Chorèmes et diagnostics de territoire : une expérience de formation. *Mappemonde*, 4 :1–6, 2002.
- A. Puissant, N. Lachiche, G. Skupinsk, A. Braud, J. Perret, et A. Mas. Classification et évolution des tissus urbains à partir de données vectorielles. *Revue Internationale de Géomatique*, 21(4) :513–532, 2011.

- D. Pumain. Villes et systèmes de villes dans l'économie. *Revue d'économie financière*, 86 : 29–46, 2006a.
- D. Pumain. *Morphogenèse. L'origine des formes*, Chapitre Systèmes de villes et niveaux d'organisation, pages 239–263. Belin, Collection Echelles, 2006b.
- D. Pumain, T. Paquot, et R. Kleinschmager. *Dictionnaire La ville et l'urbain*. Anthropos Economica, 2006, 320 p.
- K. P. Pyun, J. Lim, C. S. Won, et R. M. Gray. Image segmentation using hidden markov gauss mixture models. *IEEE Transactions on Image Processing*, 16(7) :1902–1911, 2007.
- J. Quionero-Candela, M. Sugiyama, A. Schwaighofer, et N. D. Lawrence. *Dataset Shift in Machine Learning*. The MIT Press, 2009.
- J.-B. Racine. Le modèle urbain américain : Les mots et les choses. *Annales de Géographie*, 80(440) :397–427, 1971.
- R. Raina, A. Battle, H. Lee, B. Packer, et A. Y. Ng. Self-taught learning : transfer learning from unlabeled data. In *Proceedings of the 24th Int. Conf. on Machine learning*, pages 759–766, 2007.
- C. Redfearn. The topography of metropolitan employment : Identifying centers of employment in a polycentric urban area. *Journal of Urban Economics*, 61 :519–541, 2007.
- C. Redfearn, G. Giuliano, A. Agarwal, S. He, et L. Hu. Network accessibility & the evolution of urban employment. Metrans project 06-16 draft report, USC, 2008.
- A. Rencher. *Methods of Multivariate Analysis, 2nd Edition*. Wiley Series in Probability and Statistics, 2002.
- J. Renz et B. Nebel. *Handbook of Spatial Logics*, Chapitre Qualitative Spatial Reasoning Using Constraint Calculi, pages 161–215. Springer Netherlands, 2007.
- H. Reymond. Défense et illustration d'une géographie didactique universitaire. *Espace géographique*, 25(1) :3–21, 1996.
- A. Reynaud. *Société, Espace et Justice*. Esaoce et liberté. PUF, 1981, 263 p.
- G. Ritchot. La valorisation économique de l'espace géographique. *Cahiers de géographie du Québec*, 36(98) :175–214, 1992.
- M. Roncayolo. *Lectures de villes. Formes et temps*. Parentheses, 2002.
- Y. Sahraoui, N. Lunardi, et J.-P. Antoni. Essai de typologie des espaces résidentiels à partir d'indicateurs désagrégés à l'échelle du bâtiment. Application à Besançon et à Metz. *Cybergeographie : European Journal of Geography*, 687 :25, 2014.
- L. Salvati, M. Zitti, et A. Sateriano. Changes in city vertical profile as an indicator of sprawl : Evidence from a mediterranean urban region. *Habitat International*, 38(0) :119 – 125, 2013.

- G. Saporta. *Probabilités Analyse des données et Stastique*. Edition technip, 1990, 493 p.
- L. K. Saul, K. Q. Weinberger, F. Sha, J. Ham, et D. D. Lee. *Semi-Supervised Learning*, Chapitre Spectral Methods for Dimensionality Reduction, pages 279–294. The MIT Press, 2006.
- A. Schneider et C. Woodcock. Compact, dispersed, fragmented, extensive? a comparison of urban growth in twenty-five global cities using remotely sensed data, pattern metrics and census information. *Urban Studies*, 45(3) :659–692, 2008.
- T. Schwanen, F. M. Dieleman, et M. Dijst. The impact of metropolitan structure on commute behavior in the Netherlands : a multilevel approach. *Growth and Change*, 35(3) :304–333, 2004.
- N. Schwarz. Urban form revisited - selecting indicators for characterising europeancities. *Landscape and Urban Planning*, 96 :29–47, 2010.
- J. Serra. *Image Analysis and Mathematical Morphology Vol. I*. Academic Press, 1982, 610p.
- J. Shi et J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8) :888–905, 2000.
- H. Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2) :227–244, 2000.
- M. Sigelle et R. Ronfard. Modèles de Potts et relaxation d’images de labels par champs de markov. *Traitement du signal*, 9(6) :449–458, 1992.
- A. Simon et M. Leal. Updated UMZs and corresponding methodological documentation. Technical report, European Environment Agency, 2011.
- P. Simons. *Parts : A Study in Ontology*. Clarendon Press, 1987, 404 p.
- M. Sonka, V. Hlavac, et R. Boyle. *Image Processing, Analysis, and Machine Vision (3rd Edition)*. Thomson, 2008, 829 p.
- S. Srinivasan. Quantifying spatial characteristics of cities. *Urban Studies*, 39(11) :2005–2028, 2002.
- D. Stead et S. Marshall. The relationships between urban form and travel patterns. an international review and evaluation. *European Journal of Transport and Infrastructure Research*, 1(2) :113–141, 2001.
- S. Steiniger, T. Lange, D. Burghardt, et R. Weibel. An approach for the classification of urban building structures based on discriminant analysis techniques. *Transactions in GIS*, 12(1) :31–59, 2008.
- S. Steiniger et R. Weibel. Relations among map objects in cartographic generalization. *Cartography and Geographic Information Science*, 34(3) :175–197, 2007.

- O. Stock. *Spatial and Temporal Reasoning*. Springer Science+Business Media, 1997, 402 p.
- M. Sugiyama, S. Nakajima, H. Kashima, P. von Büna, et M. Kawanabe. Direct importance estimation with model selection and its application to covariate shift adaptation. In *NIPS'07*, 2007.
- C. Tannier. Formes de villes optimales, formes de villes durables. Réflexions à partir de l'étude de la ville fractale. *Espaces et sociétés*, 3(138) :153–171, 2009.
- I. Thomas, P. Frankhauser, et D. Badariotti. Comparing the fractality of european urban neighbourhoods : do national contexts matter? *Journal of Geographical Systems*, 14(2) : 189–208, 2012.
- R. W. Thomas. *Information Statistics In Geography*. Number 31 in Concepts And Techniques in Modern Geography (CATMOG). Geo Abstracts, 1981, 42 p.
- F.-P. Tourneux. *Paris et ses franges : étalement urbain et polycentrisme*, Chapitre L'évolution de l'occupation du sol dans les franges franciliennes : des artificialisations concentrées plus qu'un étalement urbain ?, pages 101–129. Editions Universitaires de Dijon, 2006.
- D.-B. Tran, A. Piombini, M. Ignatowicz, D. Moreno, R. Frigui, A. Aguiléra, et D. Badariotti. Morphologie urbaine et mobilité dans la communauté urbaine de strasbourg. *Cybergeo : European Journal of Geography*, 702, 2014.
- Y.-H. Tsai. Quantifying urban form : Compactness versus 'sprawl'. *Urban Studies*, 42(1) : 141–161, 2005.
- M. G. Turner. Landscape ecology : What is the state of the science? *Annual Review of Ecology, Evolution, and Systematics*, 36 :319–344, 2005.
- M. Vaattovaara. Residential differentiation studies by GIS. In *Joint UNECE/EUROSTAT Work Session on Methodological Issues Involving the Integration of Statistics and Geography*, Tallinn, Estonia, 2001.
- F. van Harmelen, V. Lifschitz, et B. Porter. *Handbook of Knowledge Representation*. Elsevier, 2008, 1034 p.
- V. Vapnik. *The Nature of Statistical Learning Theory, 2nd Edition*. Spr, 2000, 314 p.
- A. C. Varzi. Parts, wholes, and part-whole relations : The prospects of mereotopology. *Modeling Parts and Whole*, 20(3) :259–286, 1996.
- F. Vella. Estimating models with sample selection bias : A survey. *The Journal of Human Resources*, 33 :127–169, 1 1998.
- J. H. von Thünen. *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationaleconomie*. Hambourg, 1826.
- M. J. Wainwright et M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2) :1–305, 2008.

- M. M. Webber. *L'urbain sans lieu ni bornes*. Editions de l'Aube, 1996, 124 p.
- C. Weber. *Images satellitaires et milieu urbain*. Hermès, 1995, 184 p.
- G. C. G. Wei et M. A. Tanner. A monte carlo implementation of the em algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, 85(411) :699–704, 1990.
- F. Wu et D. Martin. Urban expansion simulation of southeast england using population surface modelling and cellular automata. *Environment and Planning A*, 34(10) :1855–1876, 2002.
- P. Wu et T. G. Dietterich. Improving SVM accuracy by training on auxiliary data sources. In *Proceedings of the 21st Int. Conf. on Machine Learning*, 2004.
- C. Yan, N. Sang, et T. Zhang. Local entropy-based transition region extraction and thresholding. *Pattern Recognition Letters*, 24 :2935–2941, 2003.
- L. Younes. Estimation and annealing for gibbsian fields. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 24(2) :269–294, 1988.
- Y. Zahavi. Traveltime budgets and mobility in urban areas. Technical report, US Departement of transportation, 1974.
- A. Zàrraga et B. Goitisoló. Méthode factorielle pour l'analyse simultanée de tableaux de contingence. *Revue de Statistique Appliquée*, 50(2) :47–70, 2002.
- A. Zàrraga et B. Goitisoló. Simultaneous analysis and multiple factor analysis for contingency tables : Two methods for the joint study of contingency tables. *Computational Statistics & Data Analysis*, 53(8) :3171 – 3182, 2009.
- J. Zhang. The mean field theory in em procedures for markov random fields. *Signal Processing, IEEE Transactions on*, 40(10) :2570–2583, 1992.
- V. W. Zheng, E. W. Xiang, Q. Yang, et D. Shen. Transferring localization models over time. In *Proceedings of the 23rd Conference on Artificial Intelligence*, pages 1421–1426, 2008.

Modélisation multi-échelles de la morphologie urbaine à partir de données carroyées de population et de bâti

Mots-Clés Données carroyées, Structures urbaines, Classification, Transfert d'apprentissage, Raisonnement spatial

Résumé La question des liens entre forme urbaine et transport se trouve depuis une vingtaine d'années au cœur des réflexions sur la mise en place de politiques d'aménagement durable. L'essor de la diffusion de données sur grille régulière constitue dans ce cadre une nouvelle perspective pour la modélisation de structures urbaines à partir de mesures de densités affranchies de toutes les contraintes des maillages administratifs. A partir de données de densité de population et de surface bâtie disponibles à l'échelle de la France sur des grilles à mailles de 200 mètres de côté, nous proposons deux types de classifications adaptées à l'étude des pratiques de déplacement et du développement urbain : des classifications des tissus urbains et des classifications des morphotypes de développement urbain.

La construction de telles images classées se base sur une démarche de modélisation théorique et expérimentale soulevant de forts enjeux méthodologiques quant à la classification d'espaces urbains statistiquement variés. Pour nous adapter au traitement exhaustif de ces espaces, nous avons proposé une méthode de classification des tissus urbains par transfert d'apprentissage supervisé. Cette méthode utilise le formalisme des champs de Markov cachés pour prendre en compte les dépendances présentes dans ces données spatialisées. Les classifications en morphotypes sont ensuite obtenues par un enrichissement de ces premières images classées, formalisé à partir de modèles chorématiques et mis à œuvre par raisonnement spatial qualitatif.

L'analyse de ces images classées par des méthodes de raisonnement spatial quantitatif et d'analyses factorielles nous a permis de révéler la diversité morphologique de 50 aires urbaines françaises. Elle nous a permis de mettre en avant la pertinence de ces classifications pour caractériser les espaces urbains en accord avec différents enjeux d'aménagement relatifs à la densité ou à la multipolarité.

Multiscale modeling of urban morphology using gridded data

Keywords Gridded Data, Urban structures, Classification, Transfer learning, Spatial reasoning

Abstract Since a couple of decades the relationships between urban form and travel patterns are central to reflection on sustainable urban planning and transport policy. The increasing distribution of regular grid data is in this context a new perspective for modeling urban structures from measurements of density freed from the constraints of administrative division. Population density data are now available on 200 meters grids covering France. We complete these data with built area densities in order to propose two types of classified images adapted to the study of travel patterns and urban development : classifications of urban fabrics and classifications of morphotypes of urban development.

The construction of such classified images is based on theoretical and experimental which raise methodological issues regarding the classification of a statistically various urban spaces. To proceed exhaustively those spaces, we proposed a per-pixel classification method of urban fabrics by supervised transfer learning. Hidden Markov random fields are used to take into account the dependencies in the spatial data. The classifications of morphotypes are then obtained by broadening the knowledge of urban fabrics. These classifications are formalized from chorematique theoretical models and implemented by qualitative spatial reasoning.

The analysis of these classifications by methods of quantitative spatial reasoning and factor analysis allowed us to reveal the morphological diversity of 50 metropolitan areas. It highlights the relevance of these classifications to characterize urban areas in accordance with various development issues related to the density or multipolar development.

UNIVERSITÉ PARIS-EST
ÉCOLE DOCTORALE MSTIC
MATHÉMATIQUES ET SCIENCES ET TECHNOLOGIES DE
L'INFORMATION ET DE LA COMMUNICATION

THÈSE

présentée en vue d'obtenir le titre de

Docteur de l'Université Paris Est

Spécialité : SCIENCES ET TECHNOLOGIES DE L'INFORMATION
GÉOGRAPHIQUE

par

Johanna BARO

Modélisation multi-échelles de la
morphologie urbaine à partir de données
carroyées de population et de bâti

VOLUME II - Annexes

Thèse préparée à l'IFSTTAR,
au sein des laboratoires GRETTIA, DEST et LVMT

Soutenue publiquement le 25 mars 2015 devant un jury composé de :

Pierre FRANKHAUSER	Université de Franche-Comté	(Rapporteur)
Charles BOUVEYRON	Université Paris Descartes	(Rapporteur)
Michel VERLEYSSEN	Université Catholique de Louvain	(Examineur)
Christiane WEBER	Université de Strasbourg	(Examineur)
Patrice AKNIN	IFSTTAR	(Directeur)
Jean-Paul HUBERT	IFSTTAR	(Co-directeur)
Étienne CÔME	IFSTTAR	(Encadrant)
Olivier BONIN	IFSTTAR	(Encadrant)

Table des matières

A	Annexes générales	2
A.1	Sources de données exploitées	2
A.2	Échantillon des aires urbaines étudiées	7
A.3	Définitions de topologie	9
A.4	Aides à l'interprétation des analyses factorielles	11
B	Annexes cartographiques	22
B.1	Atlas des tissus urbains	22
B.2	Atlas des morphotypes	75

Annexes générales

A.1 Sources de données exploitées

Données INSEE

Données carroyées à 200 mètres Les données utilisées pour tous les travaux de classification correspondent à la diffusion de novembre 2013¹ : les données sont donc issues du RFL de 2010.

Variables utilisées :

- Identifiants géographiques du carreau habité (Id, IdINSPIRE) ;
- Nombre d'individus du carreau (Ind_c).

Variables complémentaires disponibles :

- Identifiant du rectangle (idK)
- Nombre de carreaux habités du rectangle (Nbcar)
- Nombre de ménages résidant dans le rectangle (Men)
- Surface cumulée des résidences principales, en mètres carrés (Men_surf)
- Nombre total de ménages présents depuis 5 ans ou plus dans leur logement actuel (Men_occ5)
- Nombre total de ménages en logement collectif (Men_coll)
- Nombre total de ménages de 5 personnes et plus (Men_5ind)
- Nombre total de ménages d'une personne (Men_1ind)
- Nombre total de ménages propriétaires (Men_prop)

1. http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=donnees-carroyees&page=donnees-detaillees/donnees-carroyees/donnees-carroyees-200m.htm

- Nombre total de ménages dont le revenu fiscal par u.c. est en dessous du seuil de bas revenu (Men_basr)
- Nombre total d'individus résidant dans le rectangle (Ind_r)
- Nombre total d'individus âgés de 0 à 3 ans (Ind_age1)
- Nombre total d'individus âgés de 4 à 5 ans (Ind_age2)
- Nombre total d'individus âgés de 6 à 10 ans (Ind_age3)
- Nombre total d'individus âgés de 11 à 14 ans (Ind_age4)
- Nombre total d'individus âgés de 15 à 17 ans (Ind_age5)
- Nombre total d'individus âgés de 25 ans et plus (Ind_age6)
- Nombre total d'individus âgés de 65 ans et plus (Ind_age7)
- Nombre total d'individus âgés de 75 ans et plus (Ind_age8)
- Somme des revenus fiscaux par u.c. winsorisés des individus (Ind_srf)

Zonage en aires urbaines 2010 Pour délimiter les jeux de données, le zonage en aires urbaines sur de la base communale de 2010 a été utilisé². La composition et les périmètres ont été spatialisés grâce à la base de données GEOFLA® de l'IGN³.

Données d'occupation des sols

Bases de données de l'IGN Les bases de données vectorielles de l'IGN ont pu être utilisées sur la France entière, selon les termes de la licence Enseignement et Recherche. Les objets suivants ont été utilisés soit pour créer des indicateurs sur les grilles régulières, soit en support d'habillage cartographique :

- **BD Topo®** (version 2012) :

A Réseau routier :

ROUTE_PRIMAIRE
ROUTE_SECONDAIRE
SURFACE_ROUTE

B Voies ferrées et autre : TRONCON_VOIE_FERREE

D Hydrographie : SURFACE_EAU

E Bâti :

BATI_INDIFFERENCIE
BATI_INDUSTRIEL
BATI_REMARQUABLE

- **BD Carthage**
- **GEOFLA®** (2010) : COMMUNE

2. http://www.insee.fr/fr/methodes/default.asp?page=zonages/aires_urbaines.htm

3. <http://professionnels.ign.fr/geofla>

Mode d’occupation des sols Deux BD MOS ont été utilisées pour comparer les résultats de classification des tissus urbains. Les postes d’occupations des sols utilisés sont les suivants :

- BD MOS Loire Atlantique⁴
- BdOCS CIGAL de la région Alsace⁵

Production de données

A partir des sources de données citées précédemment et des algorithmes de classifications développés dans cette thèse, deux bases de données d’indicateurs sur grilles régulières à mailles de 200 mètres de côté ont été produites :

- Une base de données d’indicateurs sur le bâti, contenant 13 variables quantitatives décrivant l’emprise surfacique des bâtiments et des emprises artificialisées qui ont été constituées par agrégation d’informations issues de la BD Topo de l’IGN (cf. tableau A.1).
- Une base de données sur l’organisation morphologique de 50 aires urbaines contenant quatre variables catégorielles (plus les 3 variables utilisées pour la classification) décrivant l’organisation de 50 aires urbaines selon différents types de tissus urbains et différents morphotypes d’organisation spatiale (cf. tableau A.2).

4. <http://data.loire-atlantique.fr/donnees/detail/modes-doccupation-des-sols-en-loire-atlantique>

5. <http://opendata.alsace.fr/accueil>

TABLE A.1 – Données carroyées à 200 mètres sur le bâti en France

Variable	Type	Description
s	double precision	surface bâtie (somme des surfaces des morceaux de bâtiments contenu dans une maille)
s_m	double precision	surface moyenne des bâtiments contenus dans une maille
v	double precision	volume total des bâtiments contenus dans une maille
h	double precision	hauteur moyenne des bâtiments contenus dans une maille
nbb	double precision	nombre de bâtiments intersectant une maille
nbb_p	double precision	nombre de bâtiments contenus dans une maille (pondération du type surface du bâtiment entier / surface du morceau de bâtiment contenu dans la maille)
std_surfmorc	double precision	écart-type de la surface des bâtiments contenus dans une maille
ent3x3	double precision	entropie de la surface des bâtiments contenus dans un voisinage de 3x3 mailles
s_lisse	double precision	surface lissée (s_lisse=0 si s=0 ou si s<surf_infra/3 ou si s<111)
surf_infra	double precision	surface des infrastructures de transport (route + fer + emprises surfaciques de transport + eau)
surf_nat	double precision	surface des espaces naturels non constructible (eau parcs naturels)
surf_arti	double precision	surface des espaces artificialisés non bâti (aire triage, surface route, réservoir, surface activité)
surf_conn	double precision	surface des cours d'eau

Référence du jeu de données : BARO Johanna, 2015, « Données carroyées à 200 mètres sur le bâti en France », <http://hdl.handle.net/TEST/10025>, Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux [Distributor] V1 [Version]

TABLE A.2 – Données carroyées à 200 mètres sur l’organisation morphologique de 50 aires urbaines

Variable	Type	Description
tu	numeric(24,15)	label de tissu urbain (sortie exact de l’algorithme par champ de Markov)
tuclean	numeric(24,15)	label de tissu urbain (application d’un lissage pour corriger les mailles grands ensembles et centre historique isolées)
tusimple	numeric(24,15)	label de tissu urbain (! version en 7 classes : lissage des éléments/régions isolées + découpage des régions d’habitat individuel groupé)
morphotype	numeric(24,15)	label de morphotype

Référence du jeu de données : BARO, Johanna, 2015, « Données carroyées à 200 mètres sur l’organisation morphologique de 50 aires urbaines en France », <http://hdl.handle.net/TEST/10026>, Institut Français des Sciences et Technologies des Transports, de l’Aménagement et des Réseaux [Distributeur] V1 [Version]

A.2 Échantillon des aires urbaines étudiées

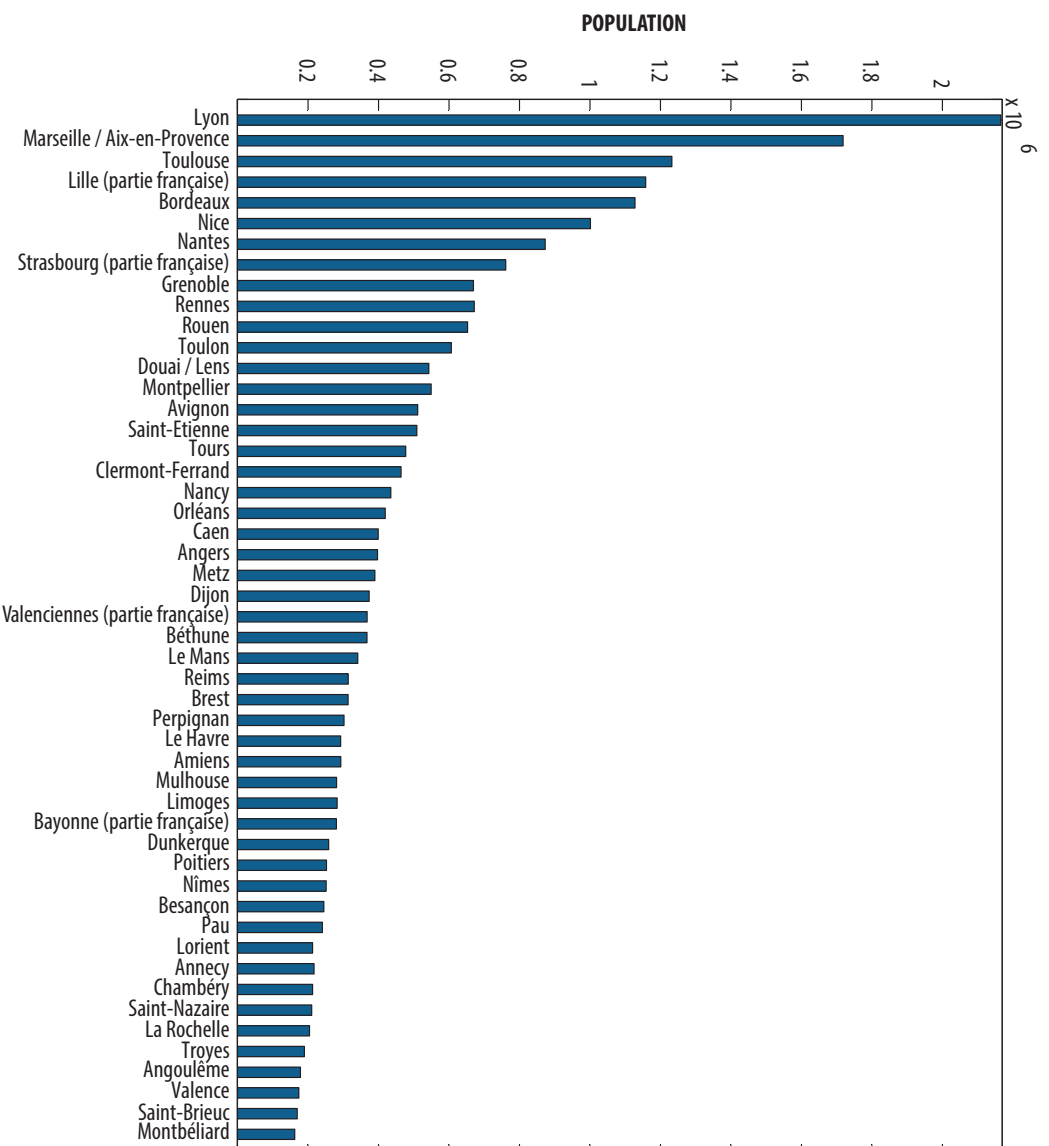


FIGURE A.1 – Population des aires urbaines

TABLE A.3 – Liste des 50 aires urbaines étudiées (zonage 2010)

Rang	Nom	Population (2010)	Nombre de commune	Surface (km ²)	Densité (hab/km ²)
2	Lyon	2165785	522	6046	358
3	Marseille - Aix-en-Provence	1718281	105	3178	541
4	Toulouse	1232398	453	5418	227
5	Lille (partie française)	1158306	125	927	1250
6	Bordeaux	1127776	255	5677	199
7	Nice	1001295	129	2594	386
8	Nantes	873133	114	3375	259
9	Strasbourg (partie française)	761042	267	2219	343
10	Grenoble	669595	196	2663	251
11	Rennes	671845	190	3778	178
12	Rouen	652898	292	2373	275
13	Toulon	606947	40	1208	502
14	Douai - Lens	542918	103	680	798
15	Montpellier	549491	115	1740	316
16	Avignon	511277	97	2081	246
17	Saint-Étienne	508847	117	1698	300
18	Tours	477438	144	3199	149
19	Clermont-Ferrand	463891	185	2435	191
20	Nancy	434948	283	2385	182
21	Orléans	419211	138	3106	135
22	Caen	399471	268	1823	219
23	Angers	397435	133	2385	167
24	Metz	389686	217	1608	242
25	Dijon	373734	295	3352	111
26	Valenciennes (partie française)	367941	90	643	572
27	Béthune	367402	122	888	414
28	Le Mans	341080	124	2027	168
29	Reims	314175	238	2393	131
30	Brest	313686	52	998	314
31	Perpignan	302185	66	1024	295
32	Le Havre	292515	81	702	417
33	Amiens	293145	256	2134	137
34	Mulhouse	281006	65	509	552
35	Limoges	282873	96	2331	121
36	Bayonne (partie française)	280538	60	1267	221
38	Dunkerque	258554	57	787	329
39	Poitiers	252381	103	2324	109
40	Nîmes	251791	50	778	324
41	Besançon	245092	263	1903	129
42	Pau	240696	167	1492	161
43	Lorient	213189	29	788	271
44	Annecy	217260	69	794	274
45	Chambéry	213251	88	828	258
46	Saint-Nazaire	210574	29	880	239
47	La Rochelle	204418	63	1015	201
48	Troyes	189811	149	2042	93
49	Angoulême	178496	108	1601	111
50	Valence	173973	41	721	241
51	Saint-Brieuc	169638	49	862	197
52	Montbéliard	162582	121	791	206

A.3 Définitions de topologie

L'ensemble des définitions suivantes est extrait du cours d'introduction à la topologie de F. Nier et D. Iftimie (Université de Rennes 1) ⁶

Définition 1 (*Espace topologique*)

On appelle espace topologique un couple (X, O) où X est un ensemble et O est une famille de parties de X , appelées ouverts, vérifiant

1. Toute réunion d'ouverts est un ouvert.
2. Une intersection finie d'ouverts est un ouvert.
3. X et \emptyset sont des ouverts.

Définition 2 (*Ouvert*)

Soit (X, d) un espace métrique. On appelle ouvert de (X, d) toute partie O de X qui vérifie

$$\forall x \in O, \exists r > 0, B(x, r) \subset O$$

$$\text{où } B(x, r) = \{y \in O, d(x, y) < r\} \quad (\text{A.1})$$

Définition 3 (*Fermé*)

Dans un espace topologique (X, O) on appelle fermé toute partie de X dont le complémentaire X^c est un ouvert.

Définition 4 (*Voisinage*)

On appelle voisinage de x dans X toute partie de X contenant un ouvert contenant x . On note $\mathcal{V}(x)$ l'ensemble des voisinages de x .

Définition 5 (*Intérieur*)

Soit A une partie de X . Un point x est intérieur à A si $A \subset \mathcal{V}(x)$. On appelle intérieur d'une partie A l'ensemble des points intérieurs à A , noté $\overset{\circ}{A}$.

Définition 6 (*Adhérence*)

Soit A une partie de X . x est adhérent à A si tout voisinage V de x contient un point de A . On appelle adhérence de A l'ensemble de points adhérents à A , noté \bar{A} .

Définition 7 (*Frontière*)

On appelle frontière de A l'ensemble $\bar{A} \cap \overline{A^c}$.

6. http://www-fourier.ujf-grenoble.fr/~demailly/L3_topologie_B/topologie_nier_iftimie.pdf

Définition 8 (*Connexité*)

Un espace topologique (X, \mathcal{T}) est connexe si les seules parties de X à la fois ouvertes et fermées sont X et \emptyset . Une partie A de X est connexe si elle est un espace connexe lorsqu'elle est munie de la topologie induite par X sur A .

A.4 Aides à l'interprétation des analyses factorielles

Analyse de la répartition de la population

TABLE A.4 – Contributions des variables à la création des axes de l'AFC

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
P1	Centre principal	0,31	0,00	0,30	0,07	0,04	0,01	0,00	0,02	0,15	0,00
P2	Centre secondaire	0,07	0,01	0,06	0,02	0,05	0,02	0,17	0,01	0,56	0,03
P3	Centre tertiaire	0,01	0,00	0,22	0,03	0,03	0,47	0,00	0,05	0,00	0,08
P4	Couronne dense	0,07	0,09	0,20	0,11	0,08	0,07	0,10	0,04	0,00	0,00
P5	Couronne peu dense	0,00	0,28	0,01	0,00	0,32	0,14	0,01	0,00	0,06	0,01
P6	Noyau connecté	0,08	0,36	0,01	0,00	0,25	0,08	0,04	0,03	0,02	0,06
P7	Noyau relais	0,07	0,02	0,06	0,01	0,03	0,02	0,01	0,48	0,02	0,25
P8	Noyau satellite	0,01	0,03	0,00	0,03	0,15	0,09	0,13	0,08	0,01	0,46
P9	Noyau isolé	0,24	0,12	0,10	0,12	0,03	0,01	0,02	0,20	0,00	0,01
P10	Grands ensembles	0,02	0,03	0,01	0,03	0,00	0,01	0,50	0,08	0,18	0,08
P11	Habitat dispersé	0,13	0,05	0,03	0,58	0,03	0,08	0,01	0,00	0,01	0,02

TABLE A.5 – Cosinus carrés entre variables initiales et composantes principales de l'AFC

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
P1	Centre principal	0,62	0,01	0,27	0,05	0,02	0,00	0,00	0,00	0,02	0,00
P2	Centre secondaire	0,35	0,03	0,14	0,03	0,07	0,02	0,12	0,00	0,23	0,01
P3	Centre tertiaire	0,06	0,01	0,42	0,04	0,04	0,38	0,00	0,02	0,00	0,02
P4	Couronne dense	0,22	0,25	0,27	0,10	0,07	0,04	0,04	0,01	0,00	0,00
P5	Couronne peu dense	0,00	0,68	0,01	0,00	0,23	0,07	0,01	0,00	0,01	0,00
P6	Noyau connecté	0,15	0,65	0,01	0,00	0,13	0,03	0,01	0,01	0,00	0,01
P7	Noyau relais	0,37	0,11	0,13	0,01	0,04	0,01	0,01	0,25	0,01	0,06
P8	Noyau satellite	0,04	0,21	0,00	0,07	0,26	0,11	0,11	0,06	0,00	0,14
P9	Noyau isolé	0,51	0,24	0,10	0,08	0,02	0,01	0,01	0,04	0,00	0,00
P10	Grands ensembles	0,10	0,17	0,04	0,07	0,00	0,02	0,43	0,06	0,09	0,02
P11	Habitat dispersé	0,32	0,11	0,03	0,47	0,02	0,04	0,00	0,00	0,00	0,00

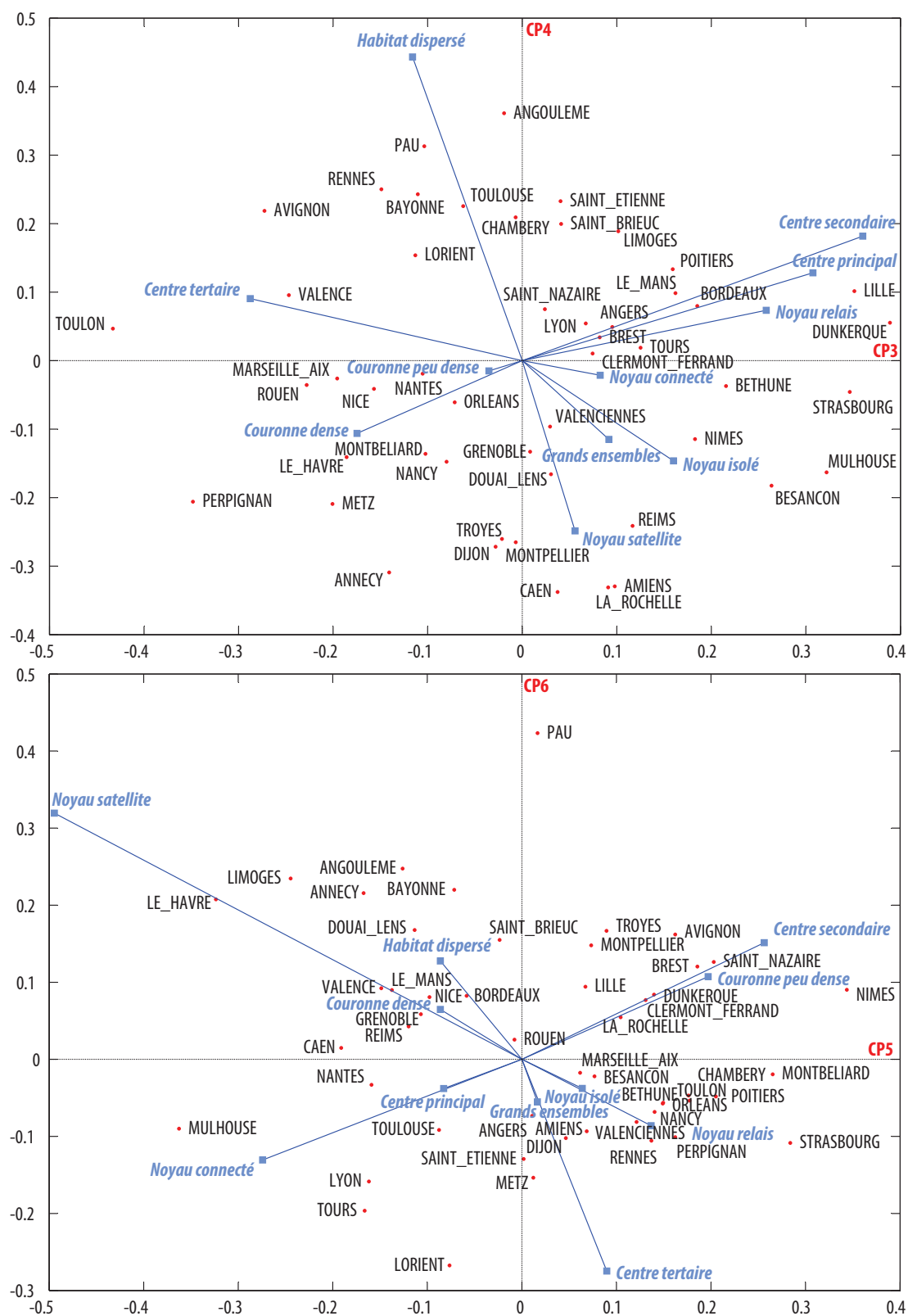


FIGURE A.2 – Deuxième et troisième plans factoriels de l'AFC

Analyse de la fragmentation intra-urbaine

TABLE A.6 – Contributions des variables à la création des axes de l'AFC

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
F1	Centre principal	0,01	0,01	0,00	0,06	0,01	0,01	0,00	0,38	0,11	0,10	0,30
F2	Centre secondaire	0,13	0,36	0,01	0,03	0,10	0,24	0,01	0,01	0,02	0,04	0,02
F3	Centre tertiaire	0,09	0,07	0,06	0,02	0,20	0,01	0,00	0,02	0,32	0,04	0,08
F4	Couronne dense	0,11	0,12	0,06	0,01	0,00	0,46	0,07	0,01	0,04	0,00	0,09
F5	Couronne peu dense	0,07	0,03	0,29	0,03	0,00	0,00	0,12	0,02	0,04	0,09	0,21
F6	Noyau connecté	0,04	0,03	0,21	0,13	0,24	0,03	0,03	0,13	0,03	0,05	0,03
F7	Noyau relais	0,01	0,01	0,01	0,04	0,15	0,01	0,00	0,06	0,22	0,44	0,02
F8	Noyau satellite	0,01	0,06	0,31	0,22	0,09	0,00	0,16	0,10	0,02	0,00	0,01
F9	Noyau isolé	0,42	0,06	0,00	0,00	0,01	0,00	0,01	0,00	0,00	0,01	0,00
F10	Grands ensembles	0,06	0,03	0,01	0,10	0,02	0,08	0,57	0,00	0,06	0,04	0,00
F11	ZA intégrée	0,04	0,00	0,01	0,11	0,00	0,16	0,04	0,26	0,10	0,01	0,23
F12	ZA périphérique	0,01	0,21	0,03	0,24	0,17	0,01	0,00	0,01	0,03	0,17	0,01

TABLE A.7 – Cosinus carrés entre variables initiales et composantes principales de l'AFC

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
F1	Centre principal	0,19	0,05	0,02	0,17	0,02	0,01	0,00	0,35	0,08	0,06	0,06
F2	Centre secondaire	0,40	0,45	0,01	0,02	0,04	0,07	0,00	0,00	0,00	0,00	0,00
F3	Centre tertiaire	0,51	0,15	0,09	0,02	0,14	0,01	0,00	0,01	0,07	0,01	0,00
F4	Couronne dense	0,47	0,21	0,07	0,01	0,00	0,20	0,02	0,00	0,01	0,00	0,00
F5	Couronne peu dense	0,38	0,05	0,43	0,03	0,00	0,00	0,05	0,01	0,01	0,02	0,01
F6	Noyau connecté	0,21	0,07	0,33	0,13	0,18	0,02	0,01	0,04	0,01	0,01	0,00
F7	Noyau relais	0,10	0,08	0,03	0,11	0,28	0,01	0,00	0,04	0,13	0,21	0,00
F8	Noyau satellite	0,05	0,13	0,45	0,21	0,06	0,00	0,06	0,03	0,00	0,00	0,00
F9	Noyau isolé	0,93	0,06	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
F10	Grands ensembles	0,40	0,08	0,01	0,12	0,02	0,05	0,30	0,00	0,02	0,01	0,00
F11	ZA intégrée	0,41	0,00	0,01	0,19	0,01	0,15	0,03	0,14	0,04	0,00	0,03
F12	ZA périphérique	0,08	0,47	0,04	0,24	0,12	0,01	0,00	0,00	0,01	0,03	0,00

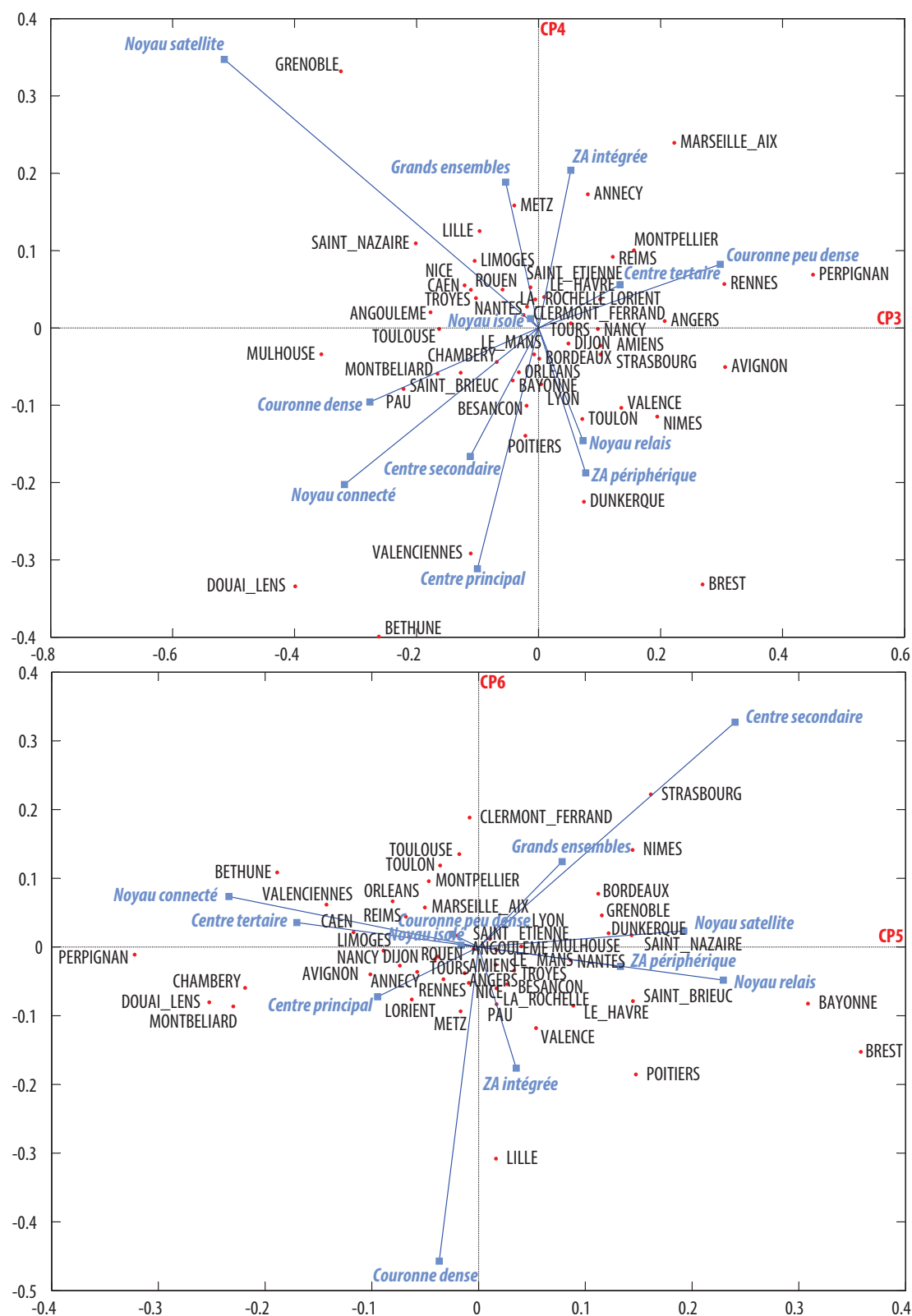


FIGURE A.3 – Deuxième et troisième plans factoriels de l'AFC

Analyse des relations et du positionnement intra-urbains

TABLE A.8 – Contributions des variables à la création des axes de l'AFC

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
CENTRE	C11/21	0,04	0,00	0,02	0,00	0,00	0,05	0,02	0,05	0,12	0,00
	C11/22	0,15	0,02	0,01	0,01	0,00	0,00	0,12	0,08	0,02	0,09
	C11/41	0,00	0,00	0,01	0,01	0,16	0,08	0,09	0,05	0,07	0,00
	C12/21	0,10	0,00	0,03	0,12	0,05	0,02	0,02	0,00	0,00	0,12
	C12/22	0,15	0,00	0,03	0,20	0,03	0,04	0,00	0,01	0,02	0,00
	C13/22	0,00	0,03	0,10	0,00	0,03	0,36	0,14	0,00	0,00	0,01
	C13/23	0,05	0,02	0,07	0,00	0,23	0,03	0,06	0,02	0,05	0,02
	C21/22	0,05	0,06	0,04	0,01	0,06	0,03	0,03	0,34	0,04	0,00
	C21/41	0,02	0,02	0,03	0,05	0,02	0,09	0,31	0,00	0,00	0,02
	C22/41	0,01	0,00	0,00	0,07	0,12	0,00	0,01	0,17	0,00	0,07
	C23/22	0,02	0,03	0,10	0,01	0,09	0,01	0,01	0,00	0,19	0,06
	C23/23	0,00	0,04	0,24	0,00	0,09	0,05	0,04	0,01	0,03	0,01
	C23/41	0,00	0,01	0,00	0,12	0,00	0,02	0,03	0,12	0,14	0,40
PÉRIPHÉRIE	P12/21	0,00	0,14	0,01	0,00	0,00	0,00	0,01	0,02	0,02	0,06
	P12/22	0,02	0,34	0,01	0,19	0,01	0,01	0,00	0,09	0,13	0,03
	P13/22	0,38	0,01	0,15	0,05	0,02	0,02	0,06	0,02	0,07	0,01
	P13/23	0,00	0,08	0,14	0,15	0,08	0,18	0,01	0,01	0,02	0,05
	P21/22	0,01	0,19	0,01	0,03	0,00	0,00	0,04	0,00	0,07	0,05

TABLE A.9 – Cosinus carrés entre variables initiales et composantes principales de l'AFC

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
CENTRE	C11/21	0,36	0,00	0,08	0,00	0,01	0,10	0,04	0,06	0,12	0,00
	C11/22	0,66	0,07	0,03	0,01	0,00	0,00	0,09	0,05	0,01	0,03
	C11/41	0,02	0,00	0,05	0,03	0,37	0,15	0,15	0,06	0,06	0,00
	C12/21	0,49	0,00	0,07	0,23	0,07	0,02	0,02	0,00	0,00	0,05
	C12/22	0,54	0,01	0,05	0,27	0,02	0,03	0,00	0,01	0,01	0,00
	C13/22	0,02	0,10	0,27	0,00	0,04	0,37	0,13	0,00	0,00	0,00
	C13/23	0,27	0,07	0,17	0,00	0,29	0,03	0,06	0,01	0,03	0,01
	C21/22	0,23	0,23	0,11	0,02	0,08	0,04	0,03	0,23	0,02	0,00
	C21/41	0,15	0,11	0,08	0,11	0,02	0,11	0,35	0,00	0,00	0,01
	C22/41	0,07	0,02	0,00	0,22	0,26	0,00	0,01	0,19	0,00	0,05
	C23/22	0,09	0,16	0,32	0,01	0,14	0,01	0,01	0,00	0,12	0,03
	C23/23	0,01	0,15	0,56	0,01	0,10	0,04	0,03	0,01	0,01	0,00
	C23/41	0,00	0,04	0,02	0,34	0,01	0,03	0,04	0,12	0,11	0,25
PÉRIPHÉRIE	P12/21	0,00	0,78	0,03	0,01	0,00	0,00	0,01	0,02	0,02	0,03
	P12/22	0,04	0,67	0,02	0,17	0,01	0,00	0,00	0,03	0,03	0,01
	P13/22	0,72	0,02	0,15	0,04	0,01	0,01	0,02	0,01	0,01	0,00
	P13/23	0,00	0,23	0,28	0,21	0,08	0,14	0,01	0,00	0,01	0,02
	P21/22	0,06	0,74	0,02	0,05	0,00	0,00	0,03	0,00	0,03	0,02



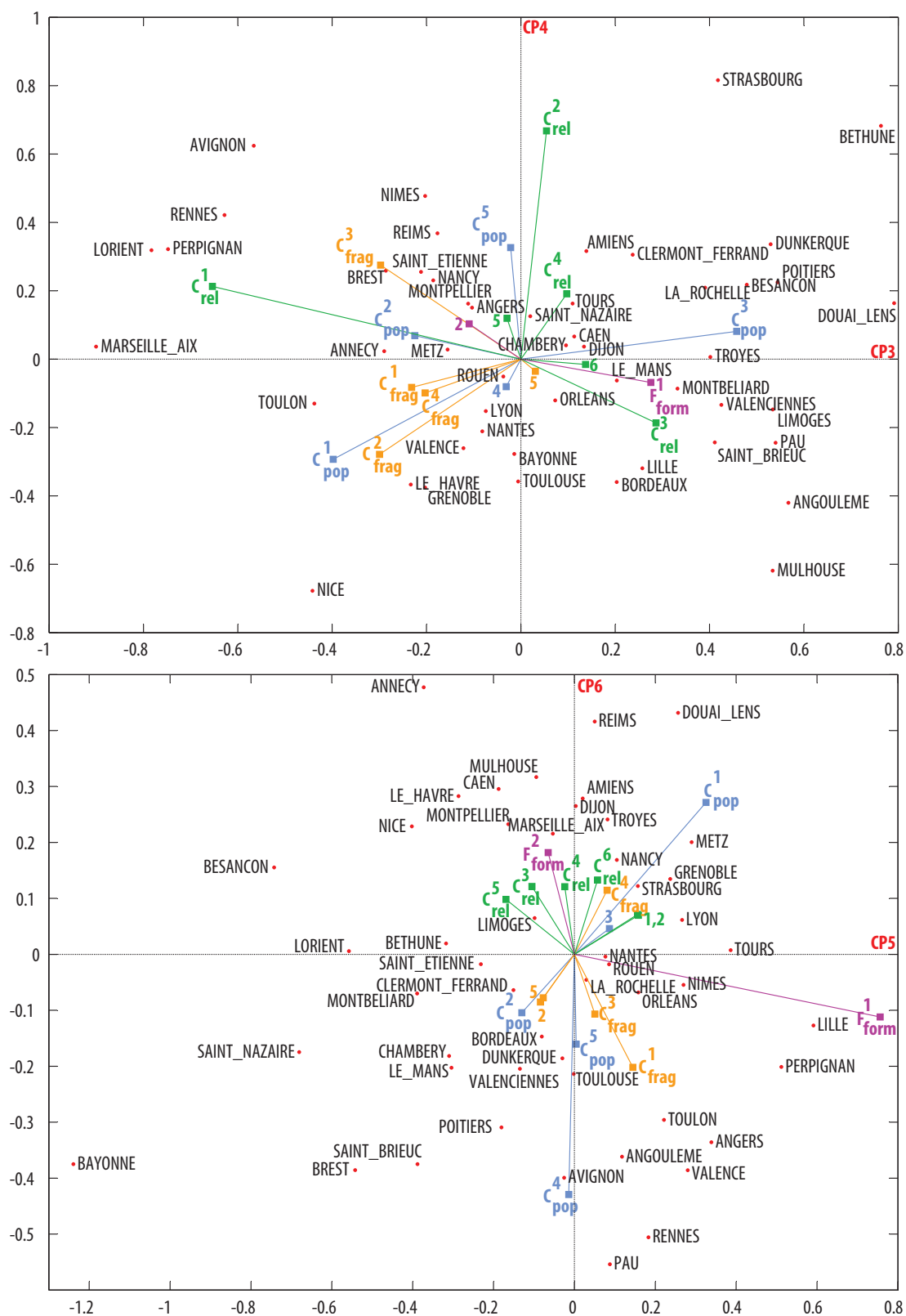
Analyse globale des caractéristiques d'organisation

TABLE A.10 – Contributions des variables à la création des axes de l'AFMM

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
POPULATION	C ¹ _{pop}	0,14	0,13	0,12	0,09	0,13	0,15	0,01	0,05	0,01	0,06
	C ² _{pop}	0,14	0,33	0,04	0,00	0,02	0,02	0,02	0,00	0,00	0,00
	C ³ _{pop}	0,00	0,04	0,16	0,01	0,01	0,00	0,09	0,10	0,04	0,09
	C ⁴ _{pop}	0,00	0,00	0,00	0,01	0,00	0,38	0,14	0,01	0,04	0,05
	C ⁵ _{pop}	0,01	0,00	0,00	0,11	0,00	0,05	0,09	0,06	0,01	0,00
FRAGMENTATION	C ¹ _{frag}	0,33	0,01	0,04	0,01	0,02	0,08	0,00	0,00	0,22	0,04
	C ² _{frag}	0,01	0,07	0,07	0,08	0,01	0,01	0,01	0,01	0,11	0,00
	C ³ _{frag}	0,00	0,01	0,07	0,08	0,00	0,02	0,01	0,00	0,03	0,07
	C ⁴ _{frag}	0,00	0,01	0,03	0,01	0,01	0,03	0,00	0,00	0,03	0,12
	C ⁵ _{frag}	0,00	0,02	0,00	0,00	0,01	0,01	0,04	0,01	0,04	0,00
RELATIONS	C ¹ _{rel}	0,09	0,16	0,32	0,05	0,03	0,01	0,05	0,01	0,00	0,00
	C ² _{rel}	0,10	0,02	0,00	0,46	0,03	0,01	0,06	0,00	0,05	0,00
	C ³ _{rel}	0,00	0,16	0,06	0,04	0,01	0,03	0,06	0,26	0,00	0,02
	C ⁴ _{rel}	0,02	0,03	0,01	0,04	0,00	0,03	0,12	0,20	0,24	0,15
	C ⁵ _{rel}	0,00	0,00	0,00	0,01	0,03	0,02	0,27	0,03	0,04	0,03
	C ⁶ _{rel}	0,00	0,00	0,01	0,00	0,00	0,04	0,00	0,24	0,04	0,05
FORME	C ¹ _{form}	0,13	0,00	0,06	0,00	0,68	0,03	0,03	0,01	0,01	0,00
	C ² _{form}	0,01	0,00	0,01	0,01	0,00	0,07	0,01	0,01	0,07	0,32

TABLE A.11 – Cosinus carrés entre composantes principales partielles et globales

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
POPULATION	C_{pop}^1	0,33	0,20	0,16	0,09	0,11	0,07	0,00	0,02	0,00	0,01
	C_{pop}^2	0,34	0,53	0,05	0,01	0,02	0,01	0,01	0,00	0,00	0,00
	C_{pop}^3	0,02	0,12	0,47	0,01	0,02	0,00	0,09	0,08	0,03	0,04
	C_{pop}^4	0,02	0,00	0,00	0,02	0,00	0,58	0,19	0,01	0,04	0,03
	C_{pop}^5	0,08	0,03	0,00	0,39	0,00	0,10	0,15	0,08	0,02	0,00
FRAGMENTATION	C_{frag}^1	0,78	0,01	0,05	0,01	0,02	0,04	0,00	0,00	0,07	0,01
	C_{frag}^2	0,08	0,28	0,23	0,20	0,02	0,02	0,01	0,01	0,08	0,00
	C_{frag}^3	0,03	0,06	0,33	0,28	0,01	0,04	0,01	0,00	0,03	0,06
	C_{frag}^4	0,01	0,12	0,23	0,05	0,04	0,07	0,00	0,00	0,05	0,14
	C_{frag}^5	0,00	0,23	0,01	0,01	0,05	0,05	0,15	0,02	0,10	0,00
RELATIONS	C_{rel}^1	0,22	0,24	0,43	0,05	0,02	0,01	0,02	0,00	0,00	0,00
	C_{rel}^2	0,30	0,03	0,00	0,56	0,03	0,01	0,03	0,00	0,02	0,00
	C_{rel}^3	0,00	0,46	0,16	0,07	0,02	0,03	0,05	0,18	0,00	0,01
	C_{rel}^4	0,12	0,11	0,02	0,10	0,00	0,04	0,13	0,18	0,19	0,08
	C_{rel}^5	0,00	0,00	0,00	0,06	0,11	0,04	0,45	0,04	0,05	0,03
	C_{rel}^6	0,02	0,02	0,09	0,00	0,02	0,09	0,00	0,42	0,06	0,05
FORME	C_{form}^1	0,31	0,00	0,08	0,00	0,57	0,01	0,01	0,00	0,00	0,00
	C_{form}^2	0,11	0,01	0,05	0,04	0,02	0,13	0,02	0,02	0,08	0,25



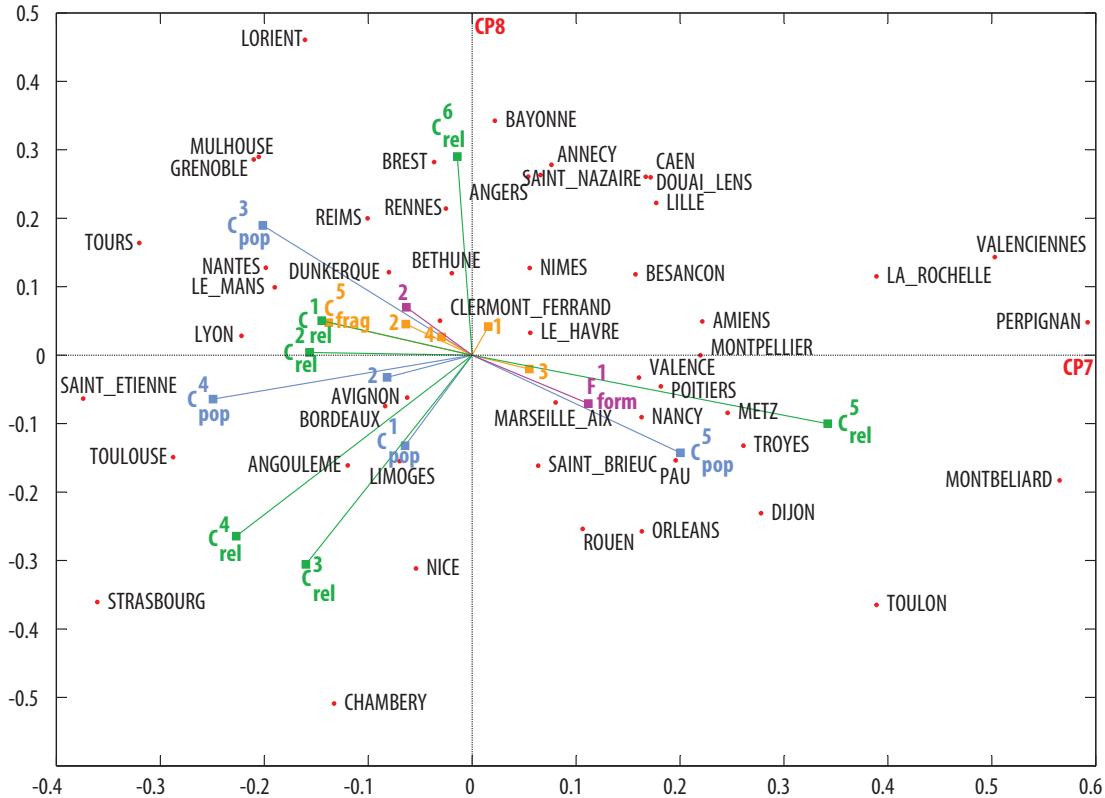


FIGURE A.5 – Plans factoriels de l’AFMM

Annexe **B**

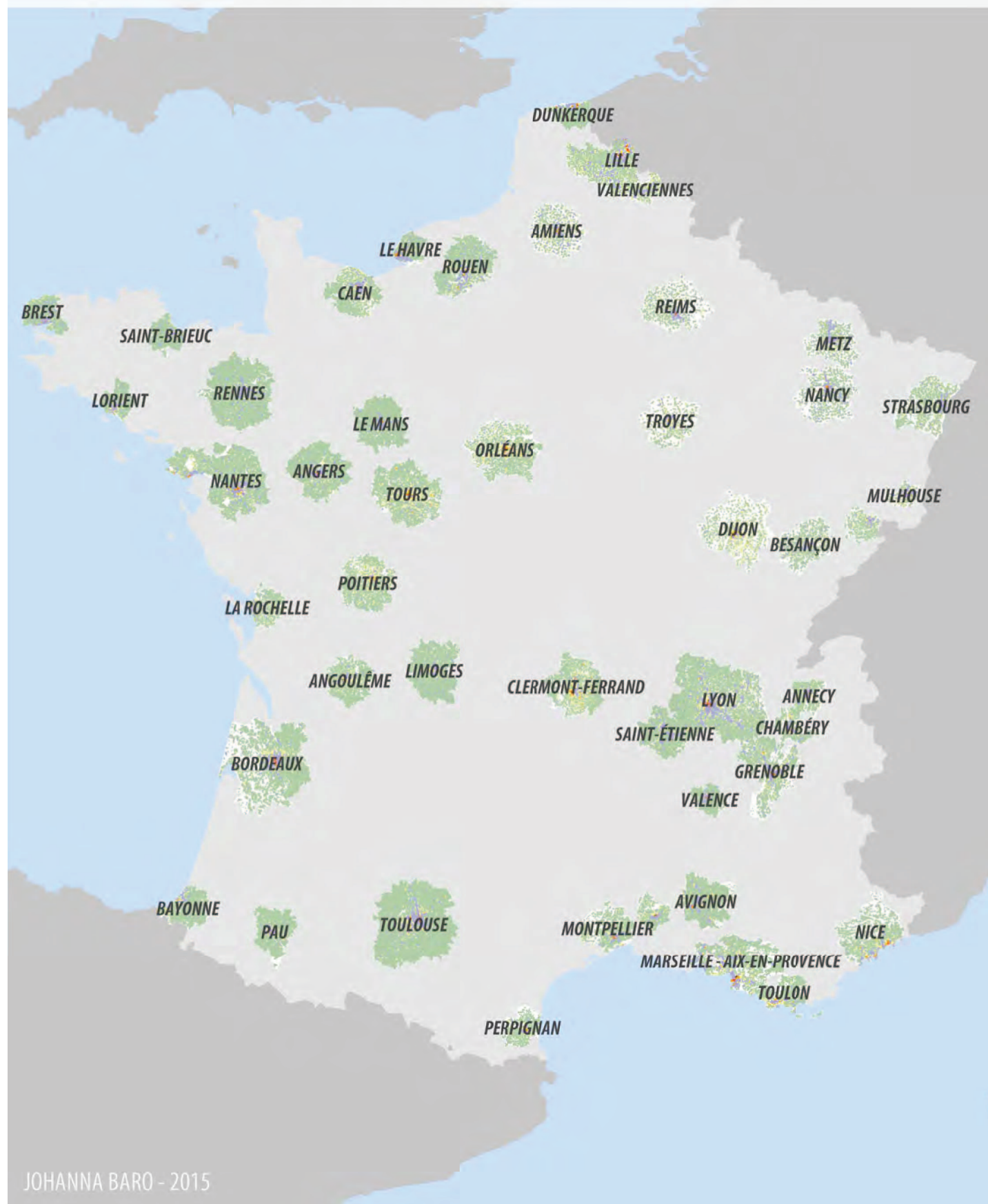
Annexes cartographiques

B.1 Atlas des tissus urbains

Référence permanent de l'atlas : BARO, Johanna, 2015, « Atlas des tissus urbains de 50 aires urbaines », <http://hdl.handle.net/TEST/10029>, Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux [Distributor]

ATLAS DES TISSUS URBAINS

DE 50 AIRES URBAINES FRANÇAISES



Atlas des tissus urbains de 50 aires urbaines françaises

Cet atlas propose une cartographie de la morphologie des cinquante plus grandes aires urbaines françaises, exception faite de Paris. Il fournit une vision de la morphologie urbaine selon une nomenclature en six catégories de tissus urbains : centre historique, habitat collectif, habitat mixte et dense, habitat individuel groupé, habitat individuel dispersé et zone d'activité.

L'atlas révèle la diversité des configurations morphologiques à partir d'une nomenclature simple de tissus urbains qui permet d'analyser la composition des aires urbaines par rapport à diverses questions d'aménagement.

Les cartes présentées ci-après correspondent aux travaux de classifications tirés de la thèse en Sciences et Technologies de l'Information Géographique de Johanna Baro, intitulée « Modélisation multi-échelles de la morphologie urbaine à partir de données carroyées de population et de bâti ».

Les classifications sont réalisées à partir de variables relatives à la population et aux surfaces bâties spatialisées sur des grilles à mailles de 200 mètres de côté. Un algorithme de classification par champ de Markov a été appliqué dans un contexte de transfert d'apprentissage pour identifier les six catégories de tissus urbains sur ces 50 aires urbaines différentes.

Source d'habillage des cartes : BD Topo®, IGN

Auteur de l'atlas : Johanna Baro, johanna.baro@gmail.com

Mai 2015

ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE D'AMIENS



0 2 4 KM

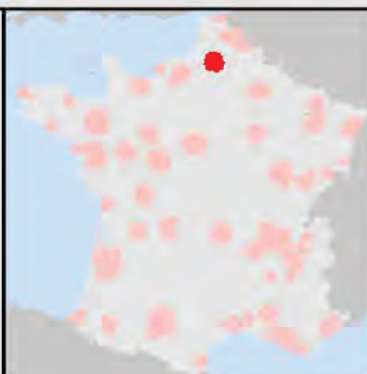
LEGENDE

TISSUS URBAINS

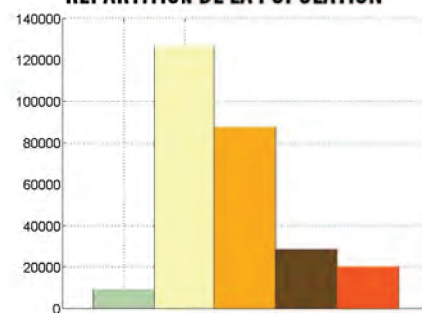
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- ⋯ Voie ferrée
- Route principale
- Surface en eau

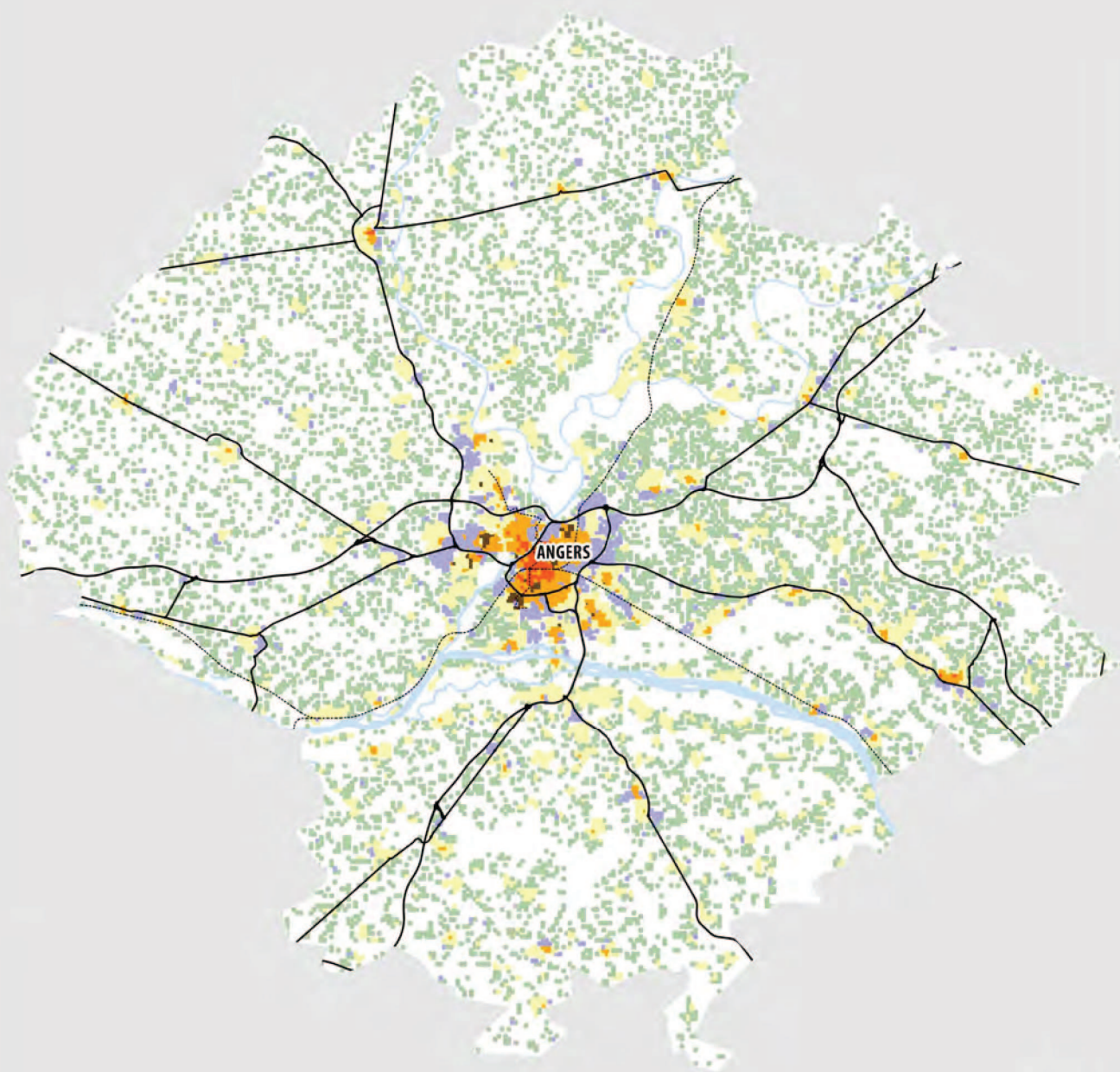


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE D'ANGERS



0 2 4 KM



LEGENDE

TISSUS URBAINS

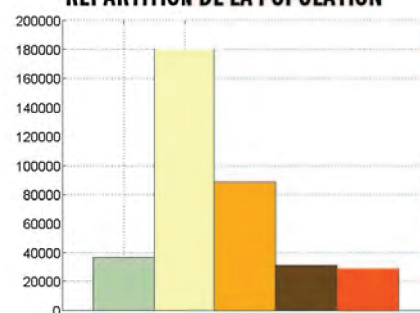
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau

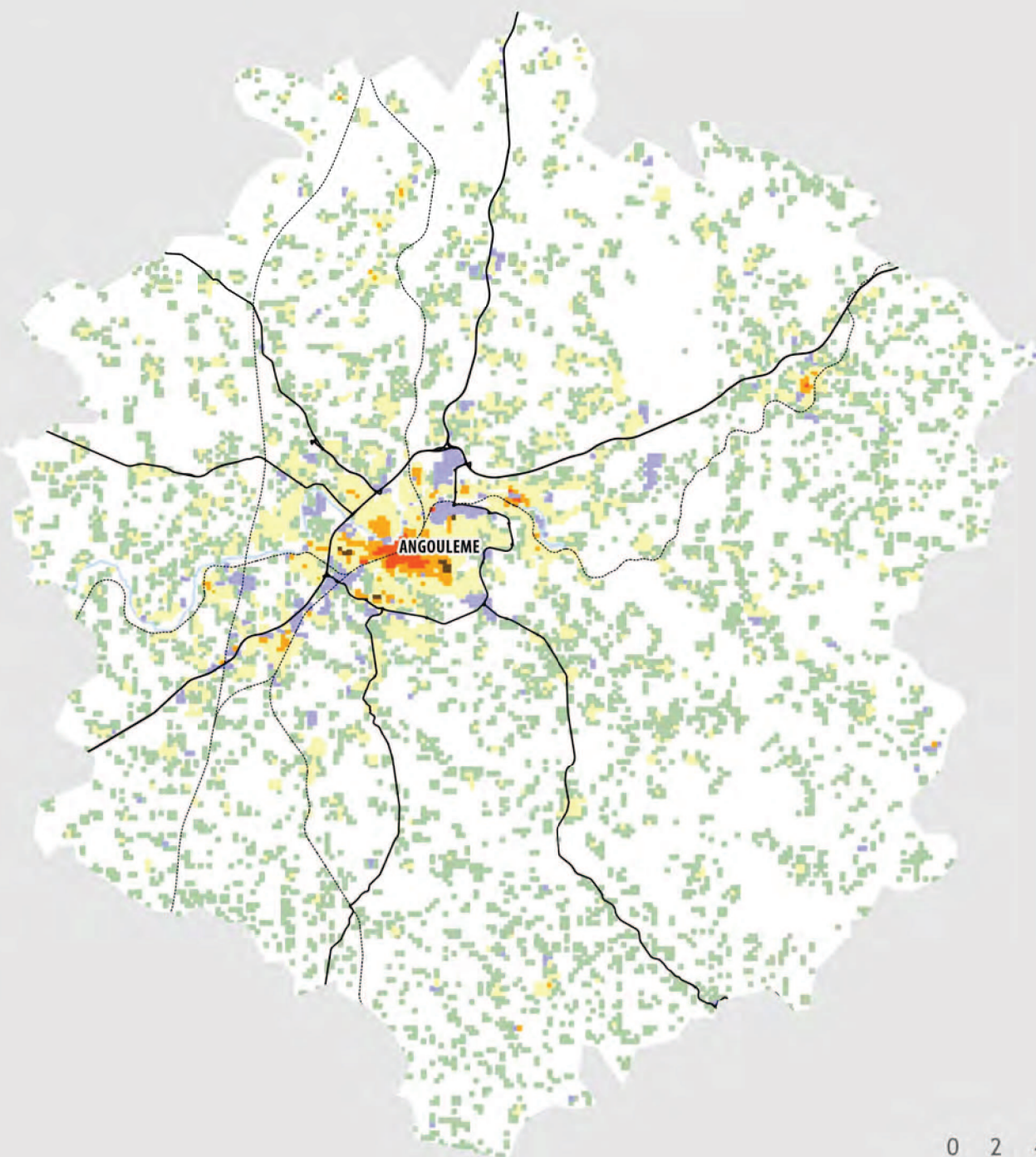


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE D'ANGOULÊME



LEGENDE

TISSUS URBAINS

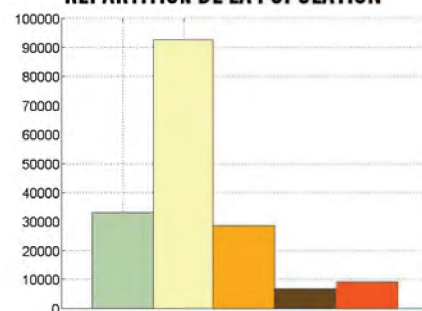
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

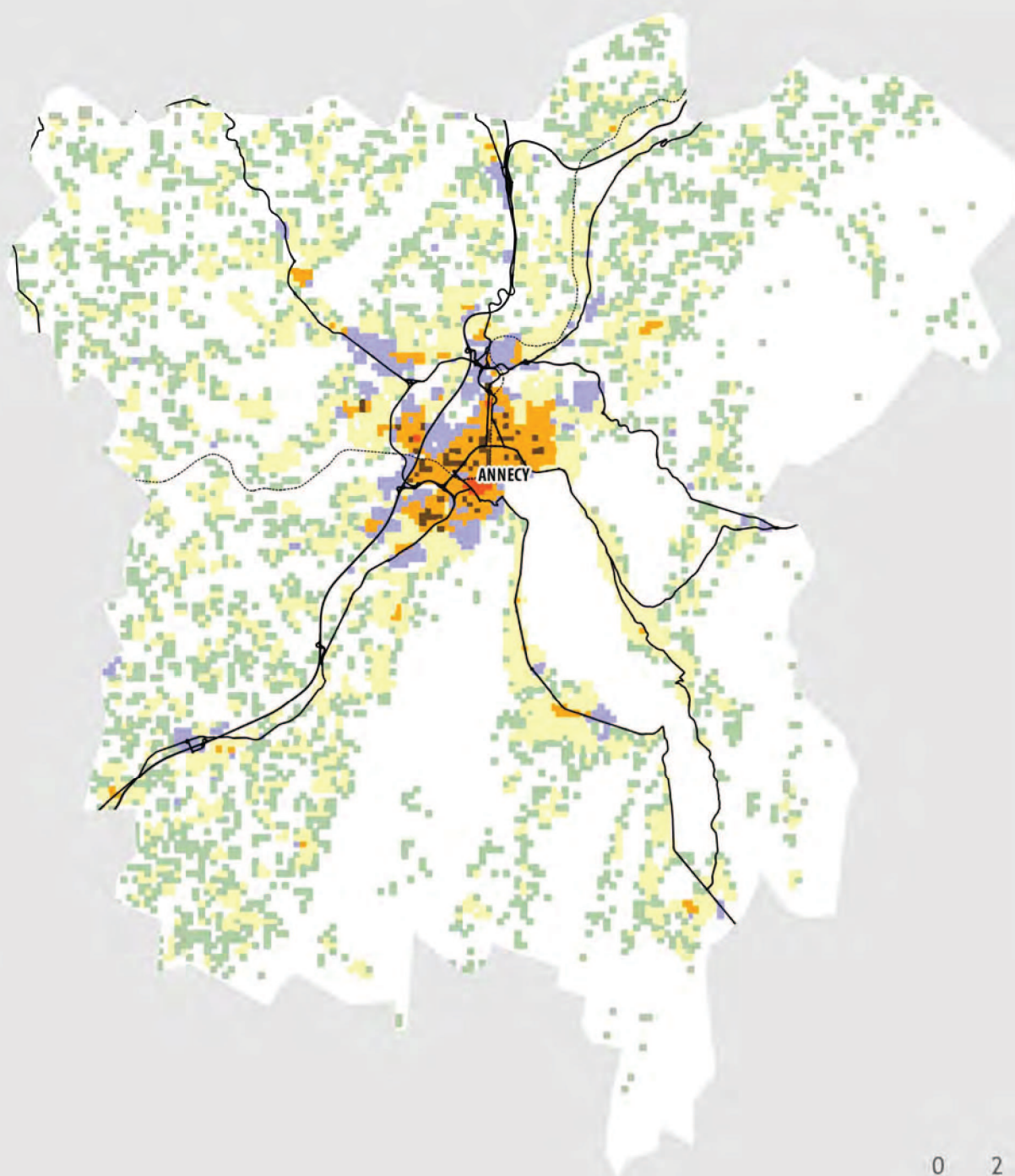


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE D'ANNECY



LEGENDE

TISSUS URBAINS

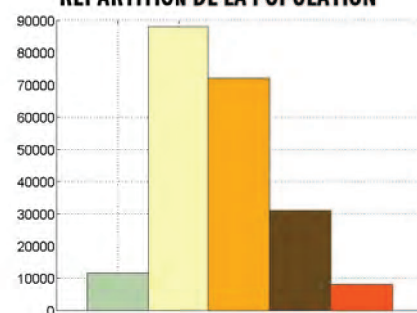
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE D'AVIGNON



0 2 4 KM



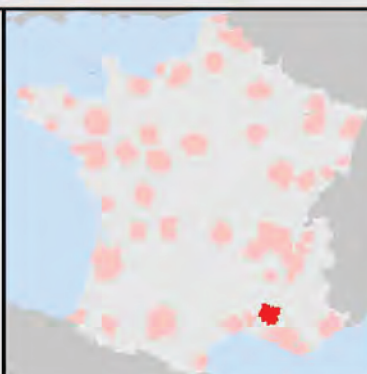
LEGENDE

TISSUS URBAINS

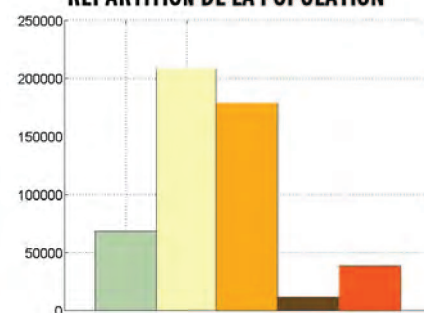
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

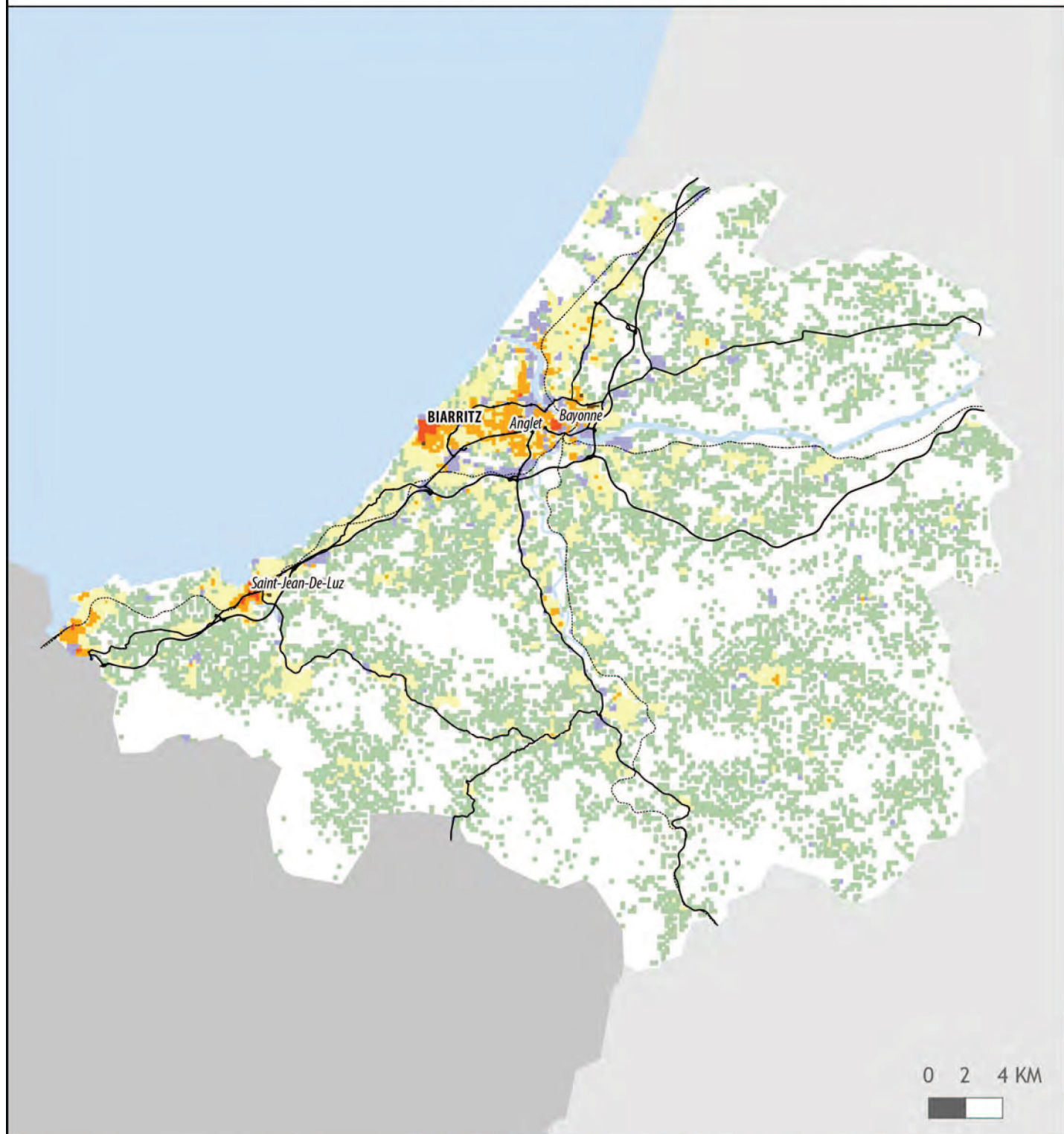


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE BAYONNE (PARTIE FRANÇAISE)



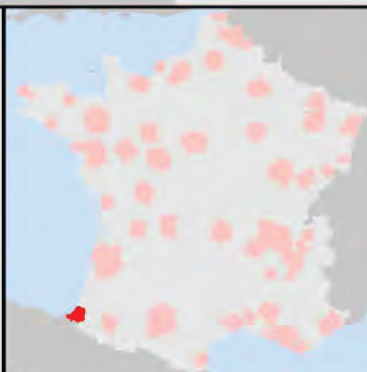
LEGENDE

TISSUS URBAINS

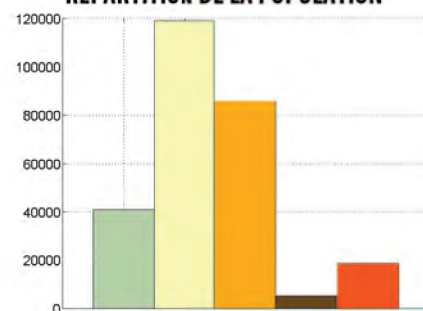
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE BESANCON



0 2 4 KM



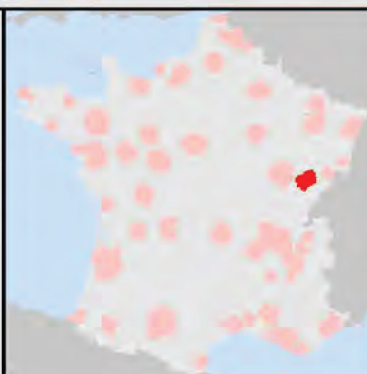
LEGENDE

TISSUS URBAINS

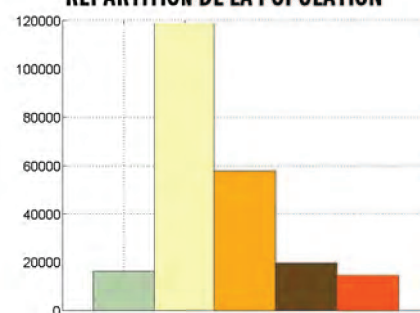
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- ⋯ Voie ferrée
- Route principale
- Surface en eau

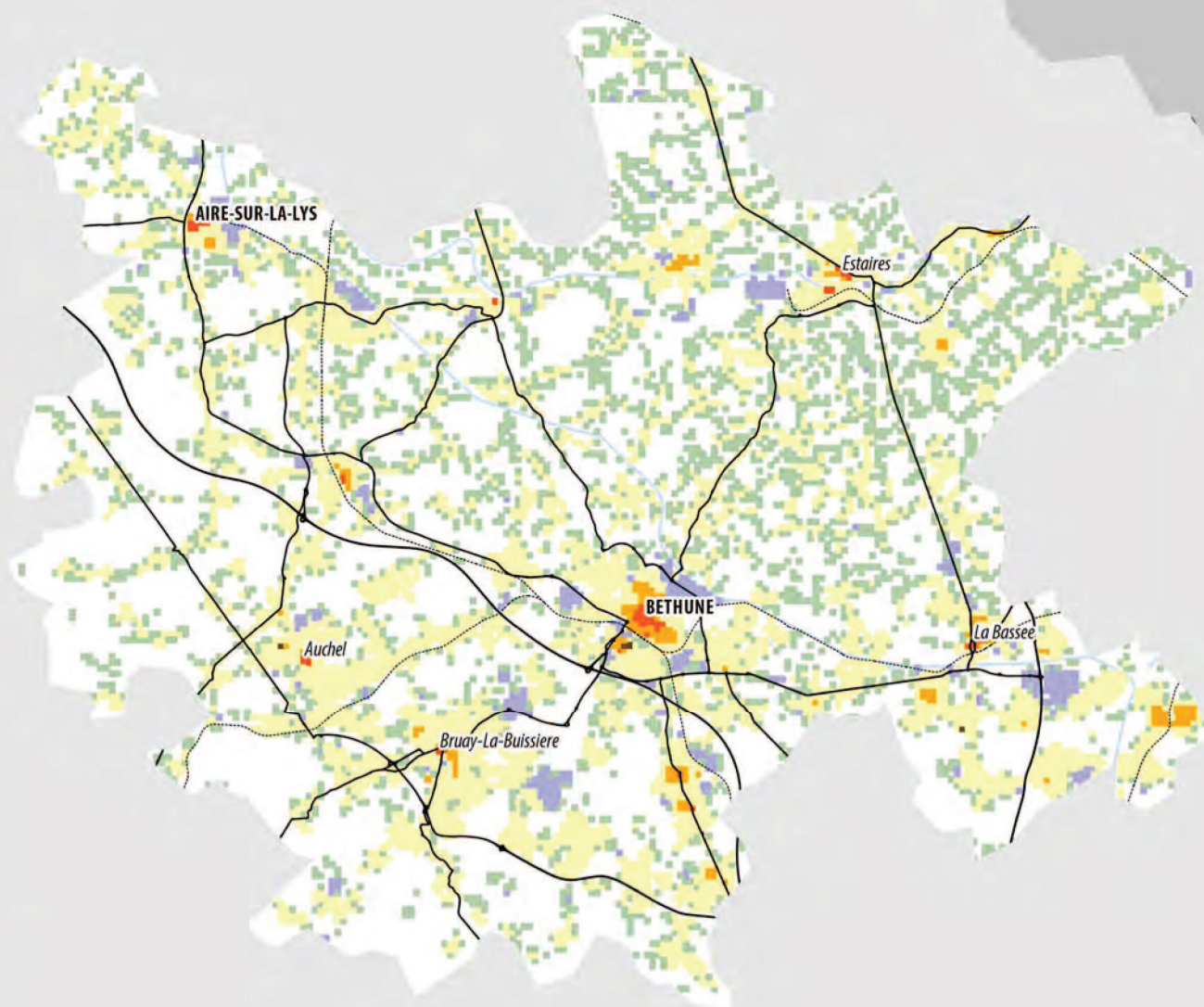


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE BÉTHUNE



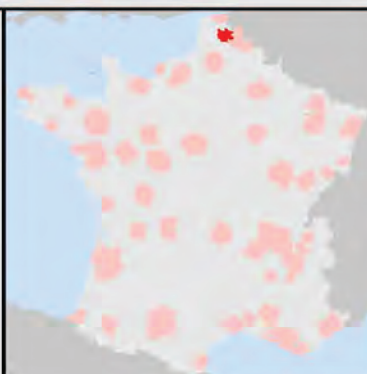
LEGENDE

TISSUS URBAINS

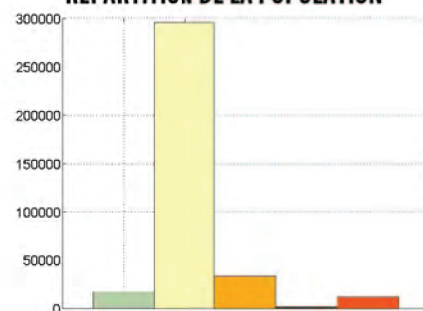
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

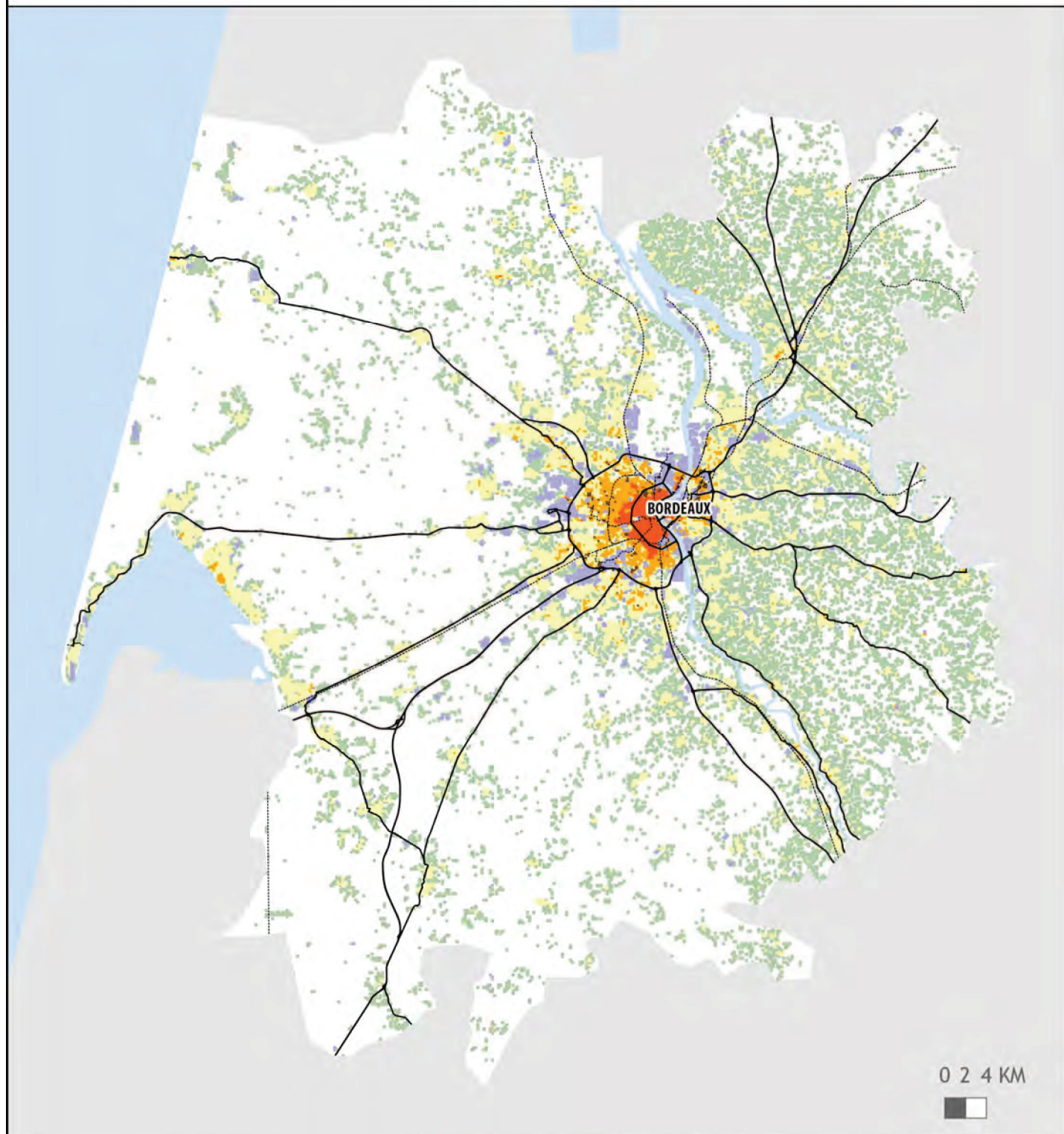


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE BORDEAUX



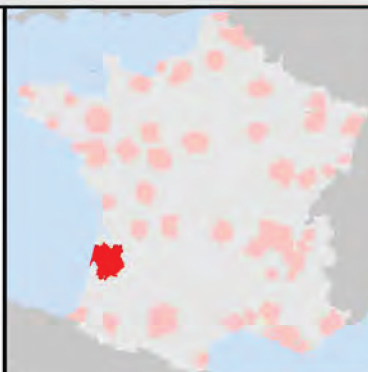
LEGENDE

TISSUS URBAINS

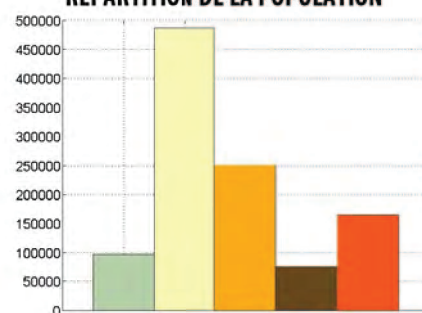
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

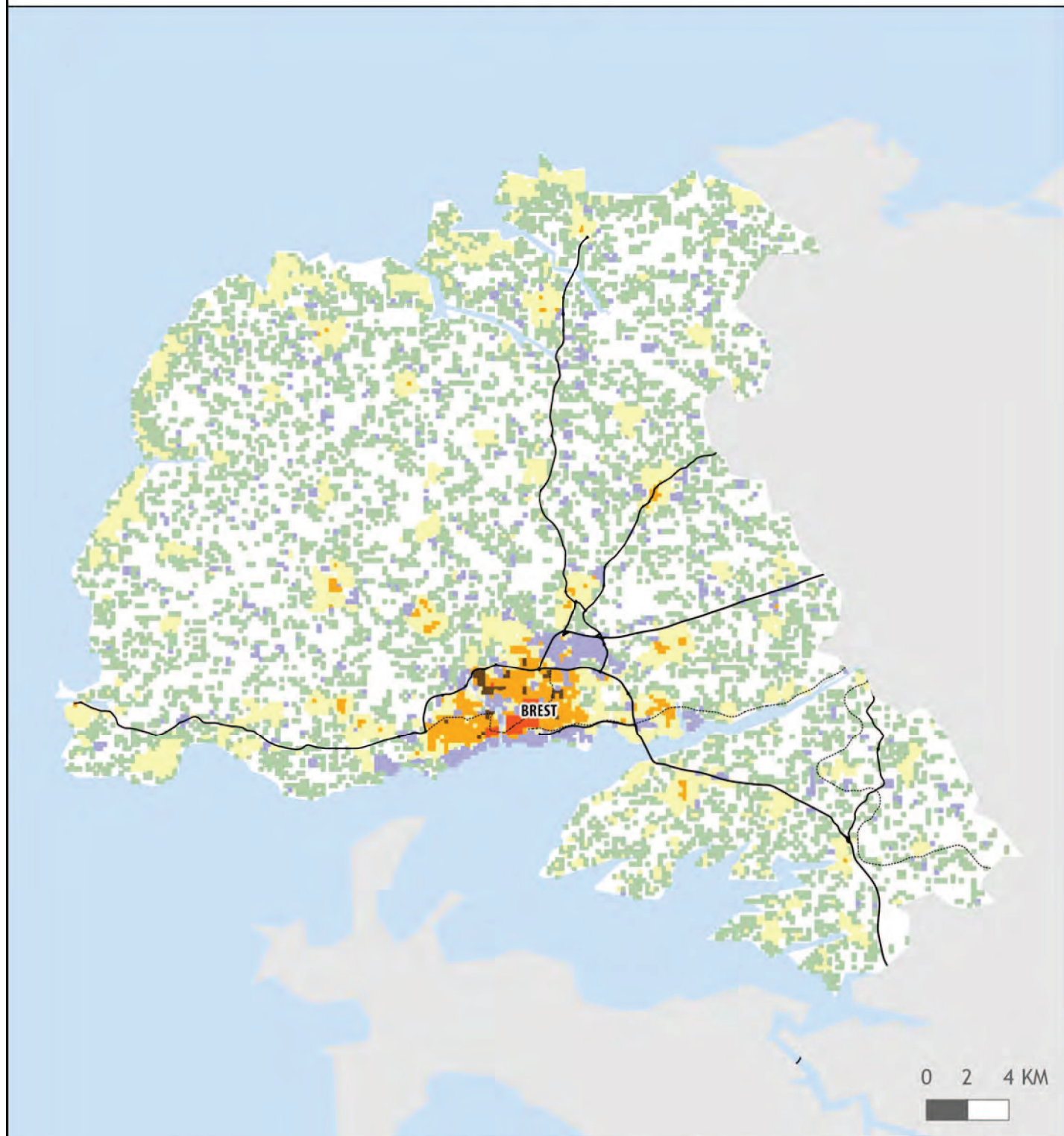


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE BREST



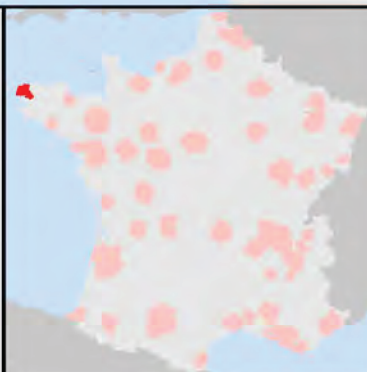
LEGENDE

TISSUS URBAINS

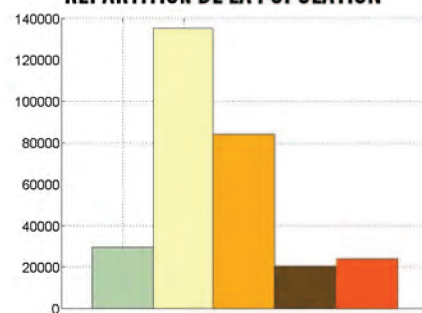
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- ⋯ Voie ferrée
- Route principale
- Surface en eau

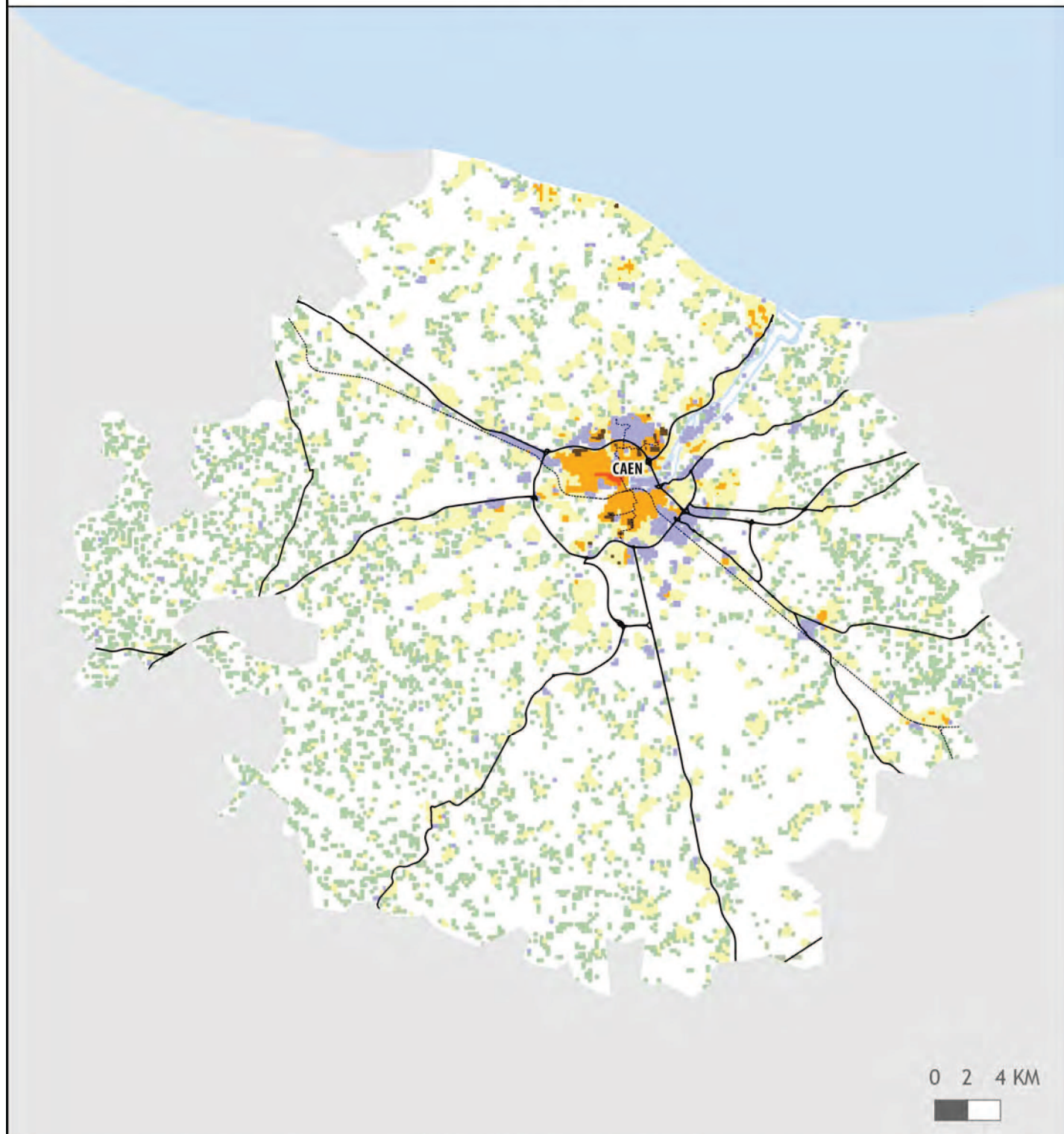


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE CAEN



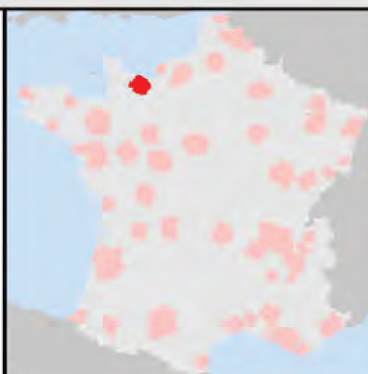
LEGENDE

TISSUS URBAINS

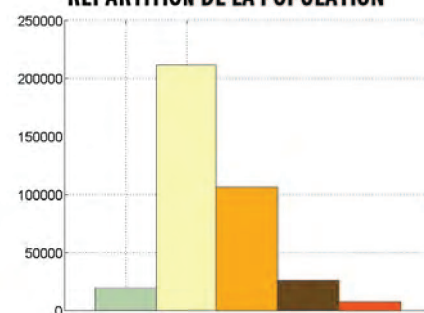
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - - - Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE CHAMBERY



0 2 4 KM



LEGENDE

TISSUS URBAINS

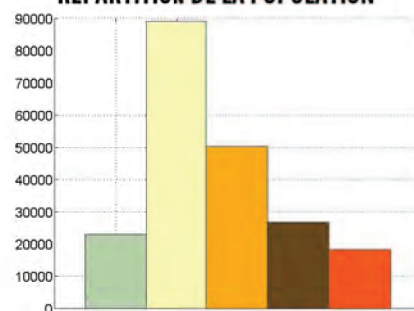
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

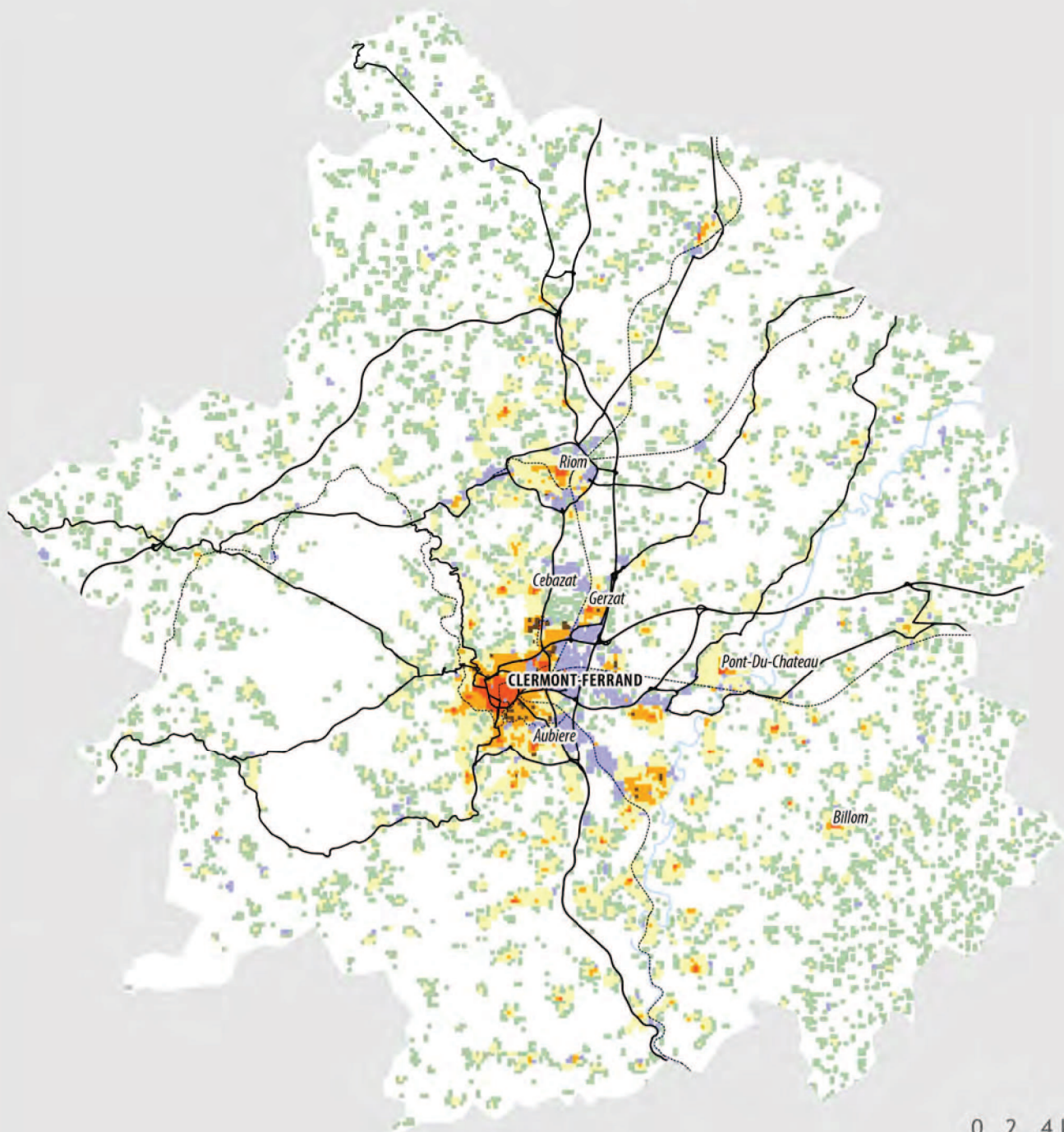


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE CLERMONT-FERRAND



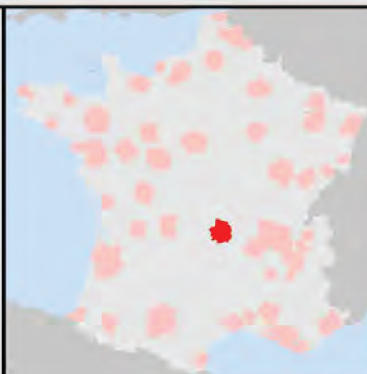
LEGENDE

TISSUS URBAINS

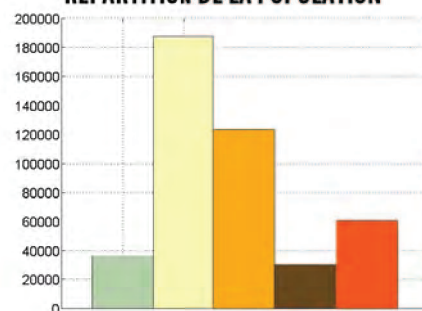
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- · — · — Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE DIJON



LEGENDE

TISSUS URBAINS

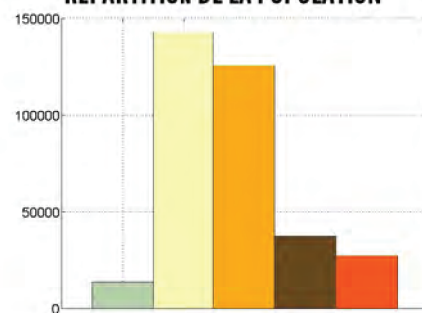
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

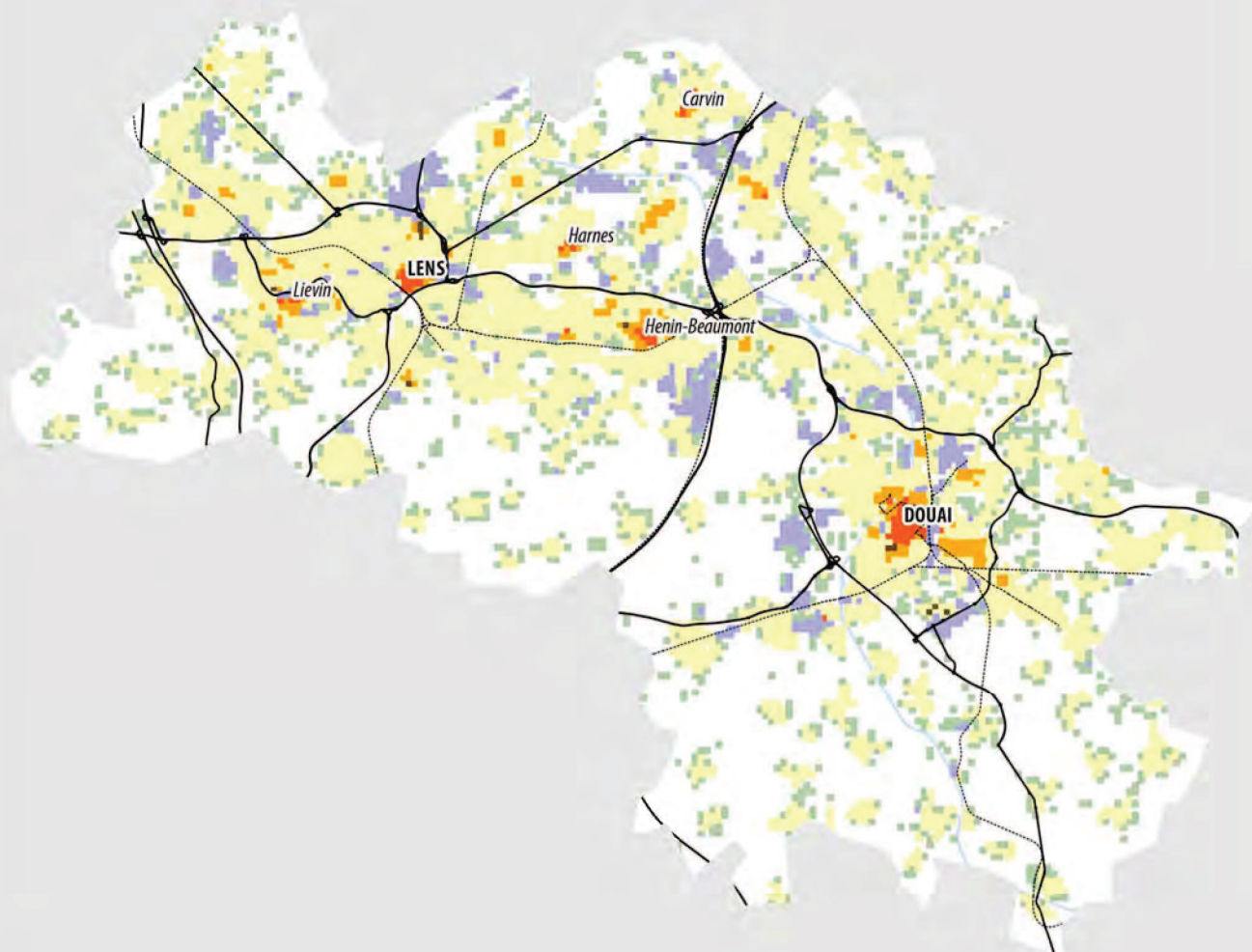


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE DOUAI - LENS



0 2 4 KM

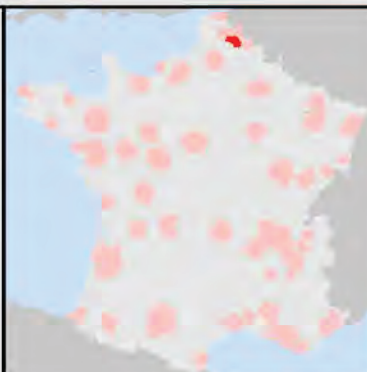
LEGENDE

TISSUS URBAINS

- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

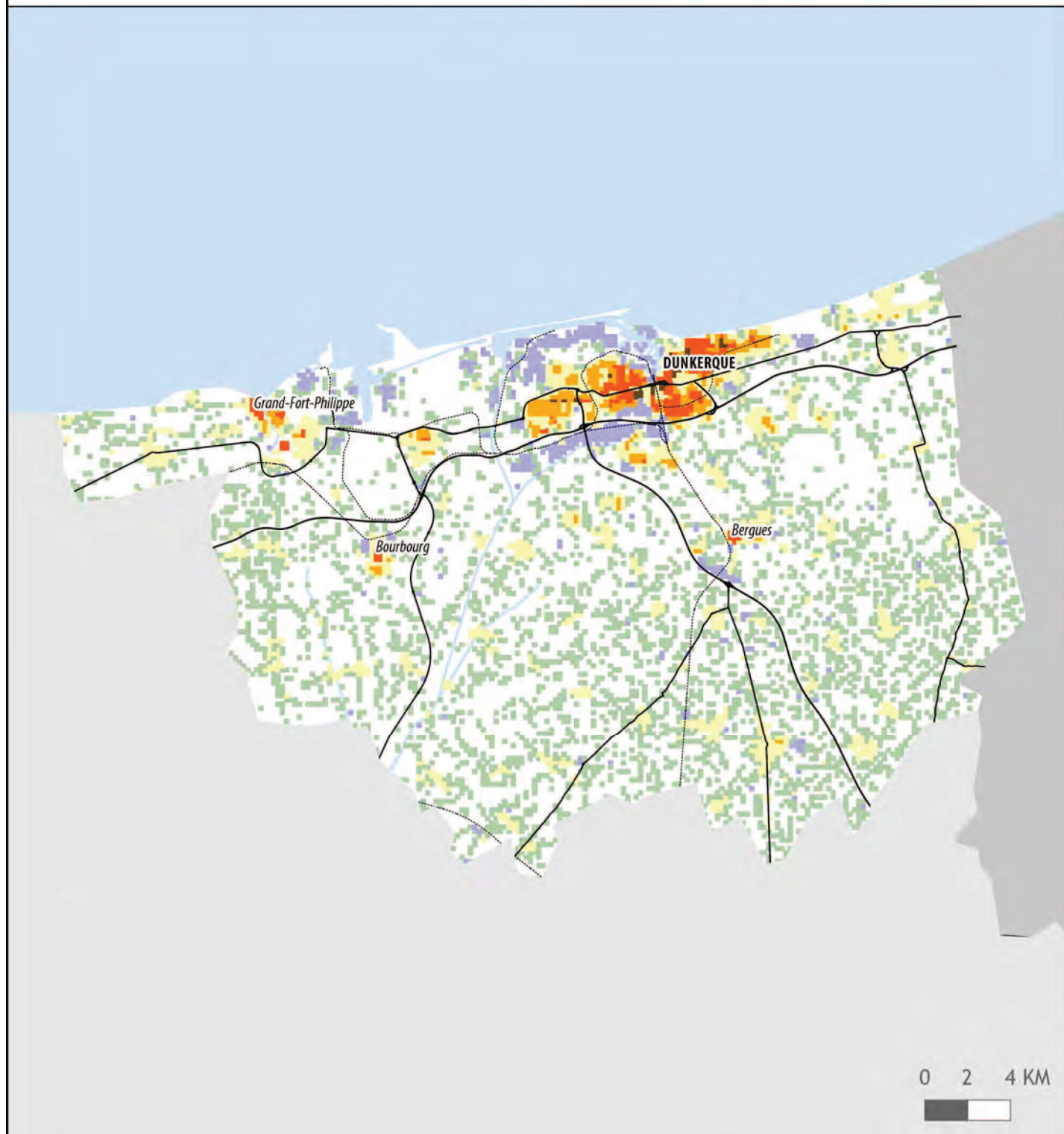


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE DUNKERQUE



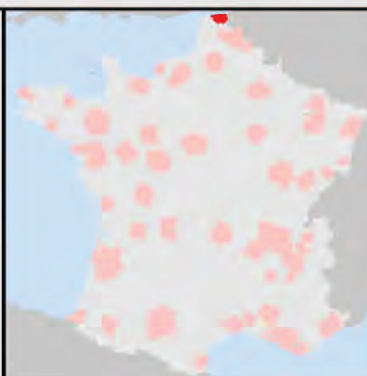
LEGENDE

TISSUS URBAINS

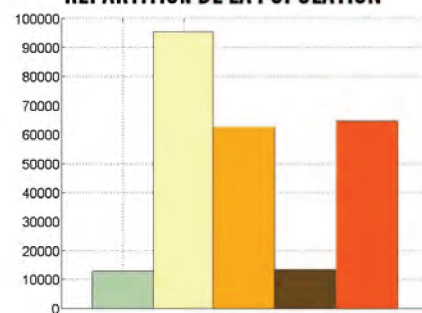
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau

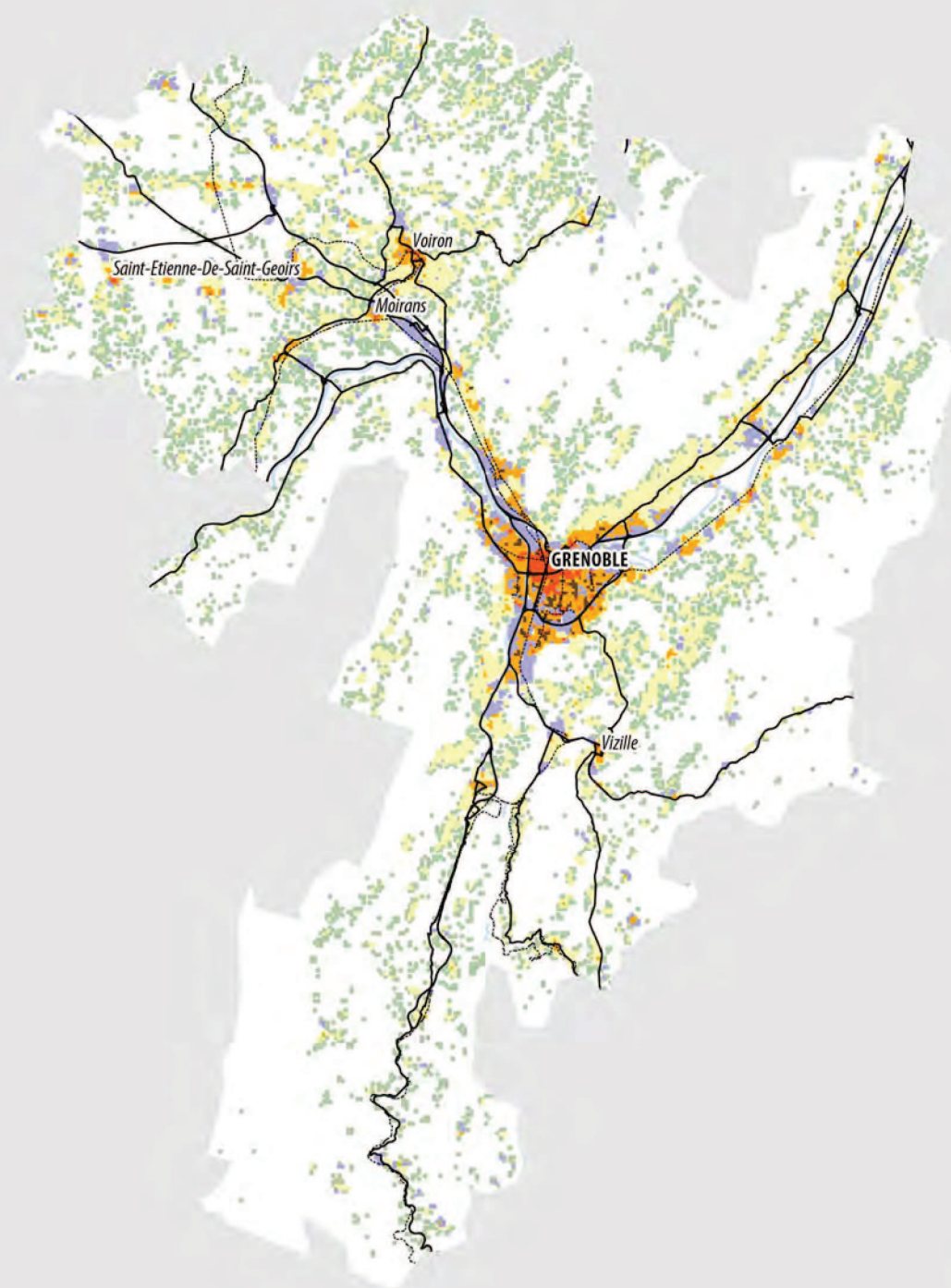


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE GRENOBLE



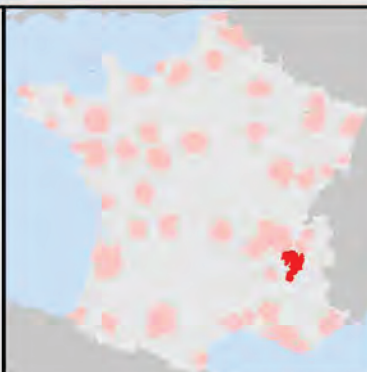
LEGENDE

TISSUS URBAINS

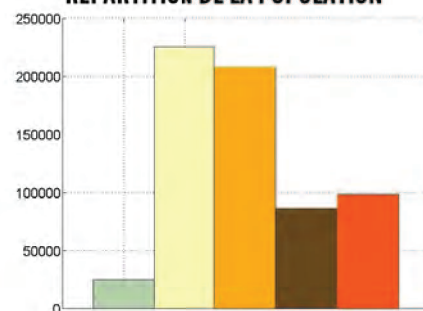
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

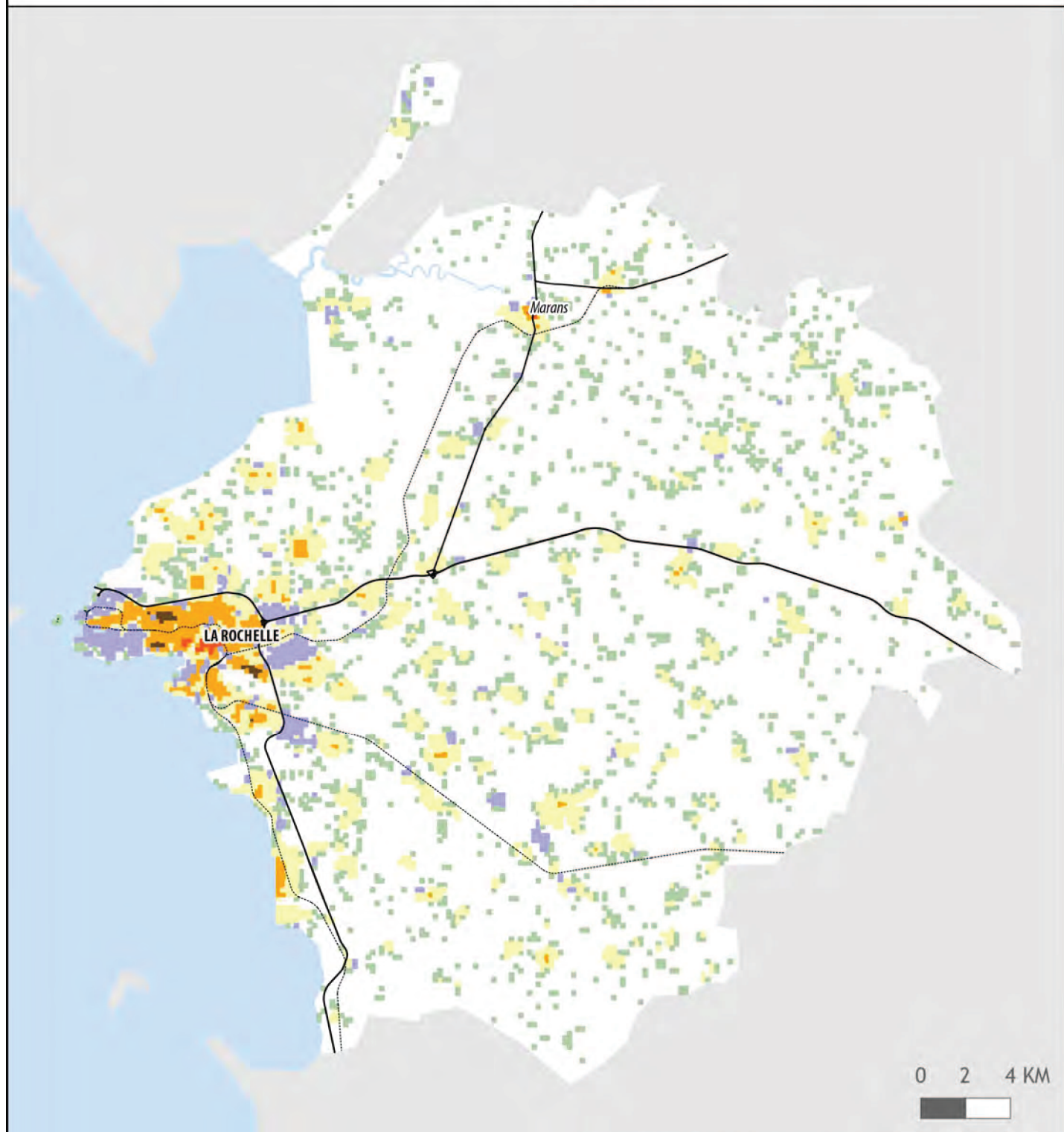


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE LA ROCHELLE



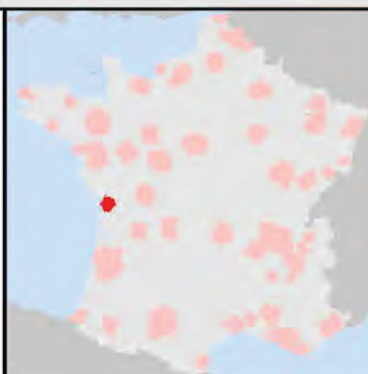
LEGENDE

TISSUS URBAINS

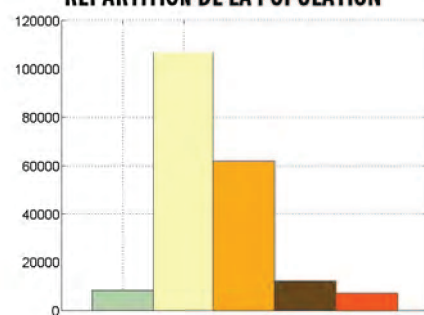
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- ⋯ Voie ferrée
- Route principale
- Surface en eau

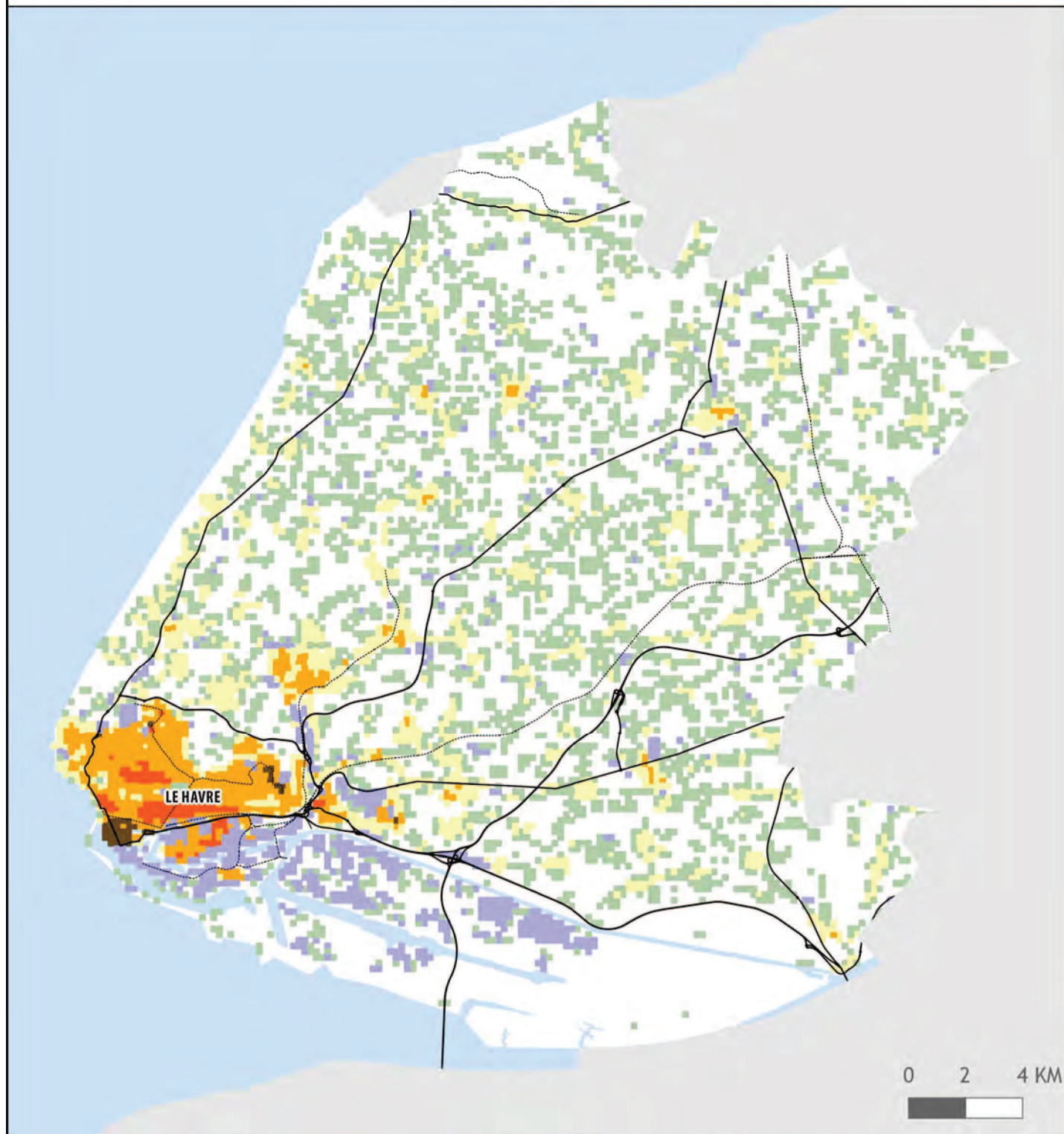


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DU HAVRE



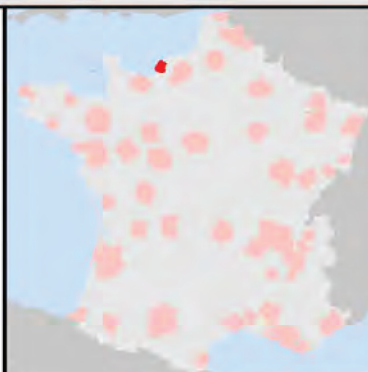
LEGENDE

TISSUS URBAINS

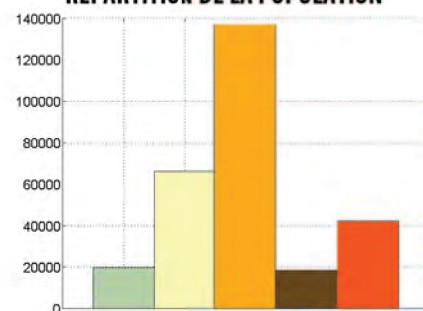
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DU MANS



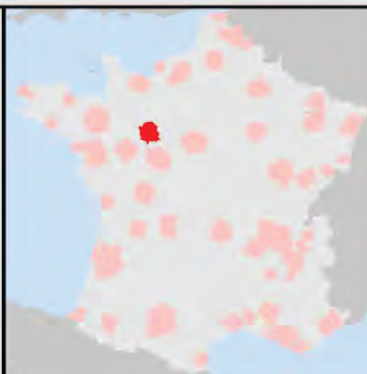
LEGENDE

TISSUS URBAINS

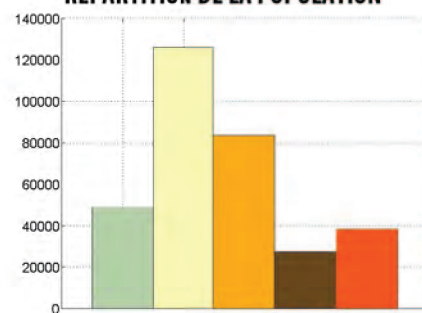
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

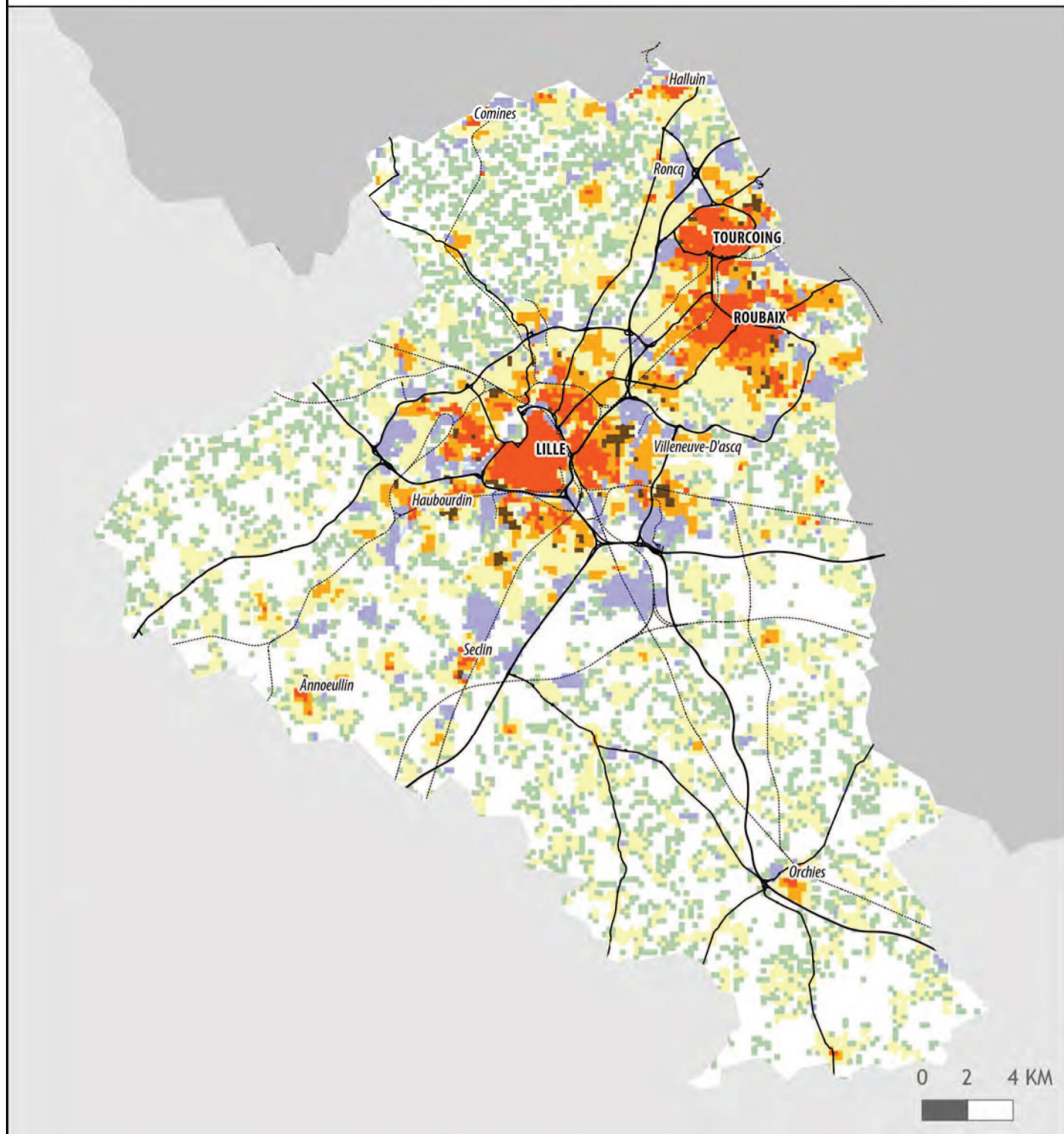


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE LILLE (PARTIE FRANÇAISE)



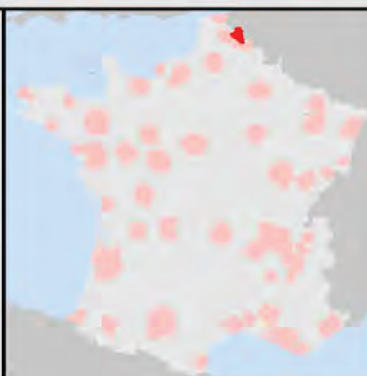
LEGENDE

TISSUS URBAINS

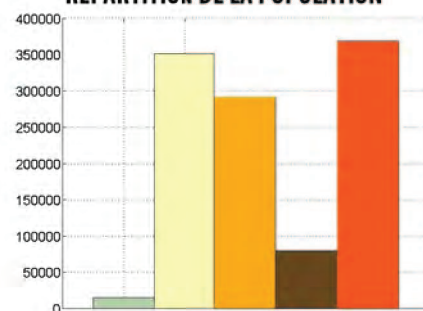
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

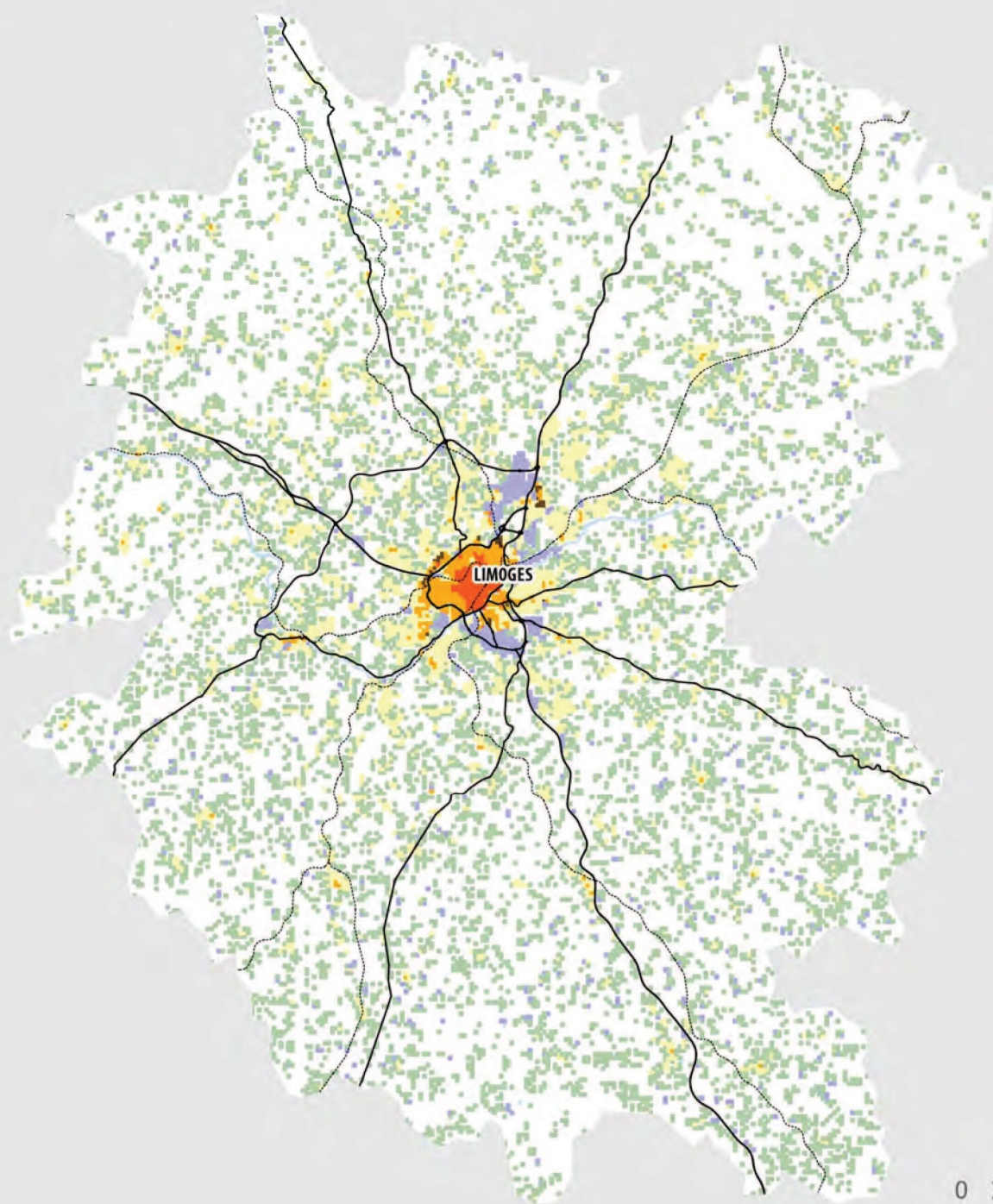


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE LIMOGES



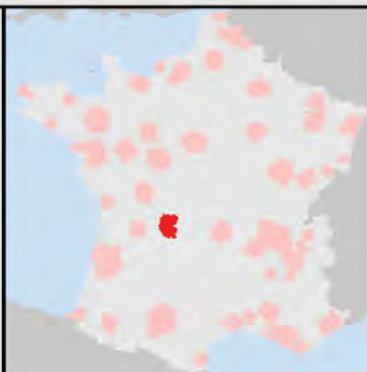
LEGENDE

TISSUS URBAINS

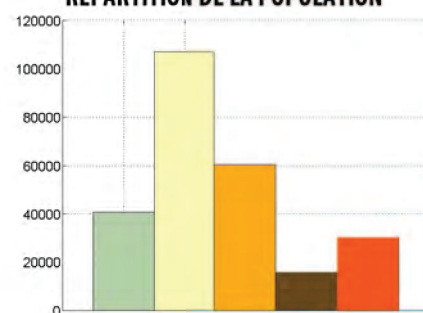
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau

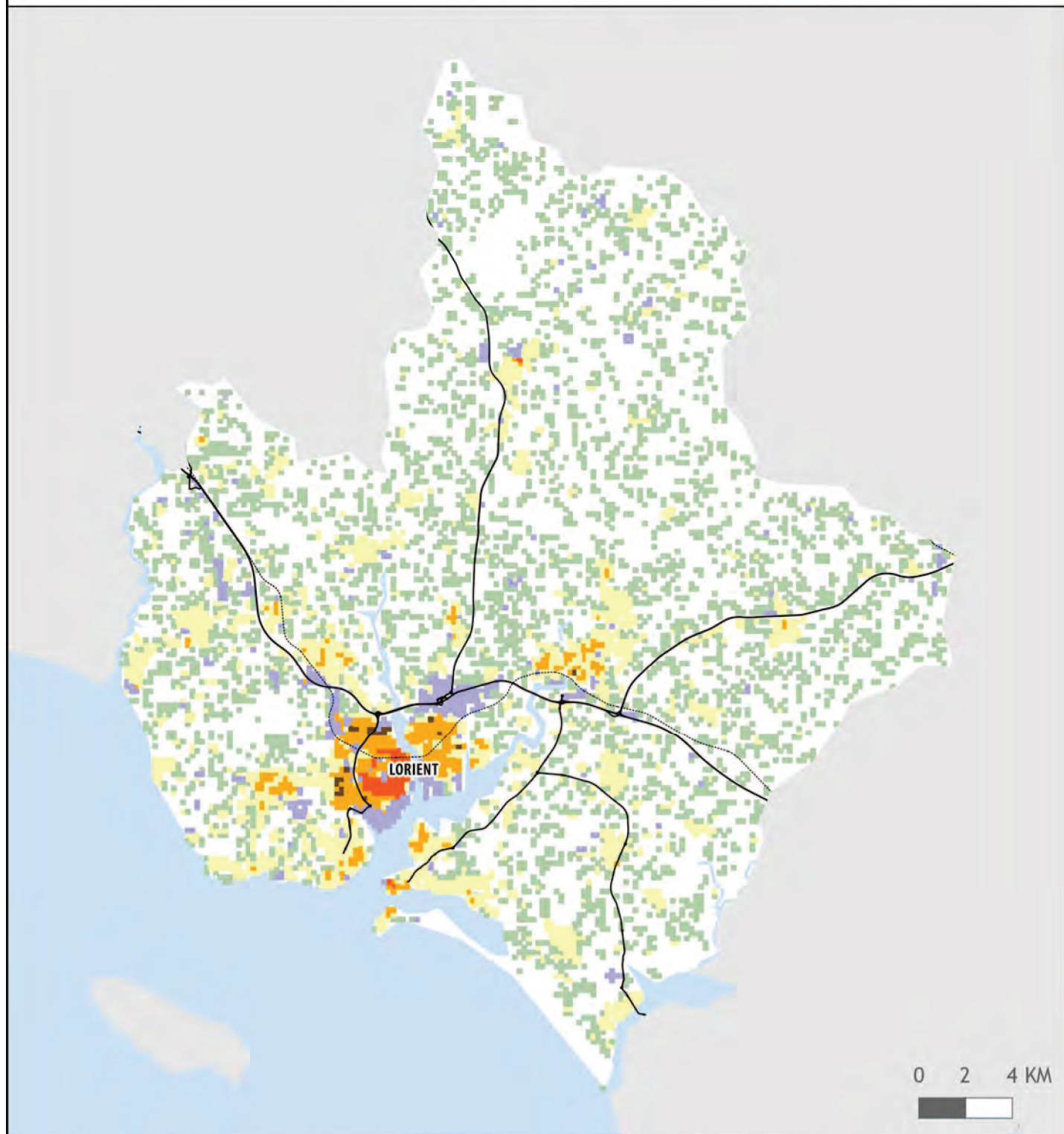


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE LORIENT



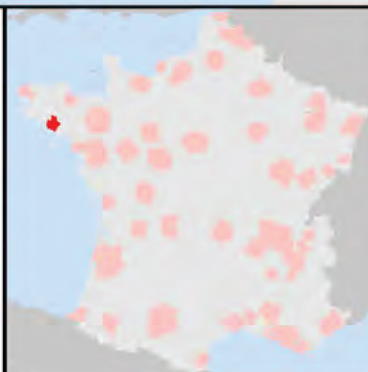
LEGENDE

TISSUS URBAINS

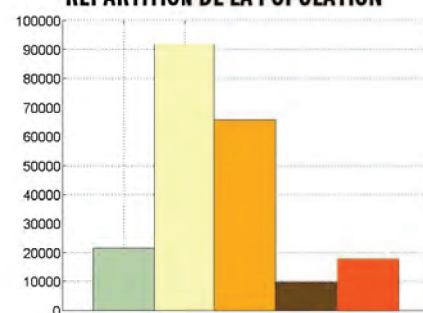
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- ⋯ Voie ferrée
- Route principale
- Surface en eau

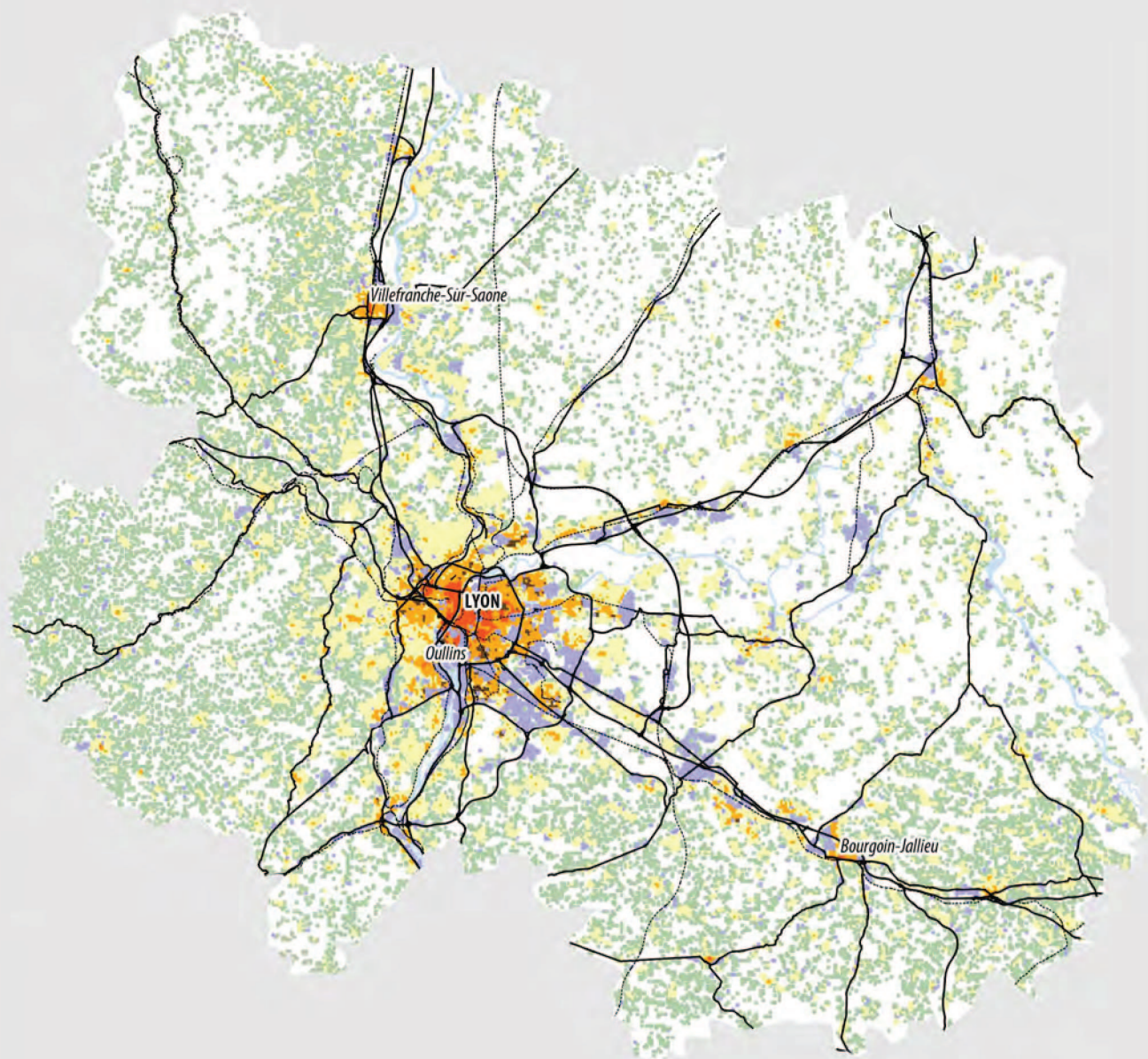


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE LYON



0 2 4 KM

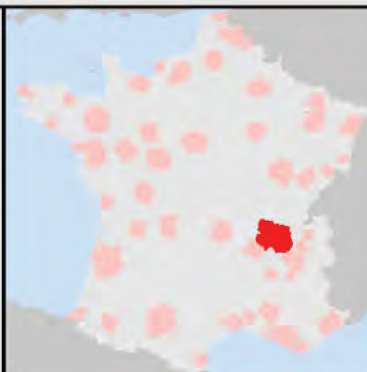
LEGENDE

TISSUS URBAINS

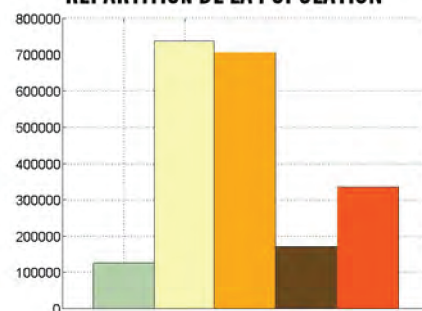
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- ⋯ Voie ferrée
- Route principale
- Surface en eau

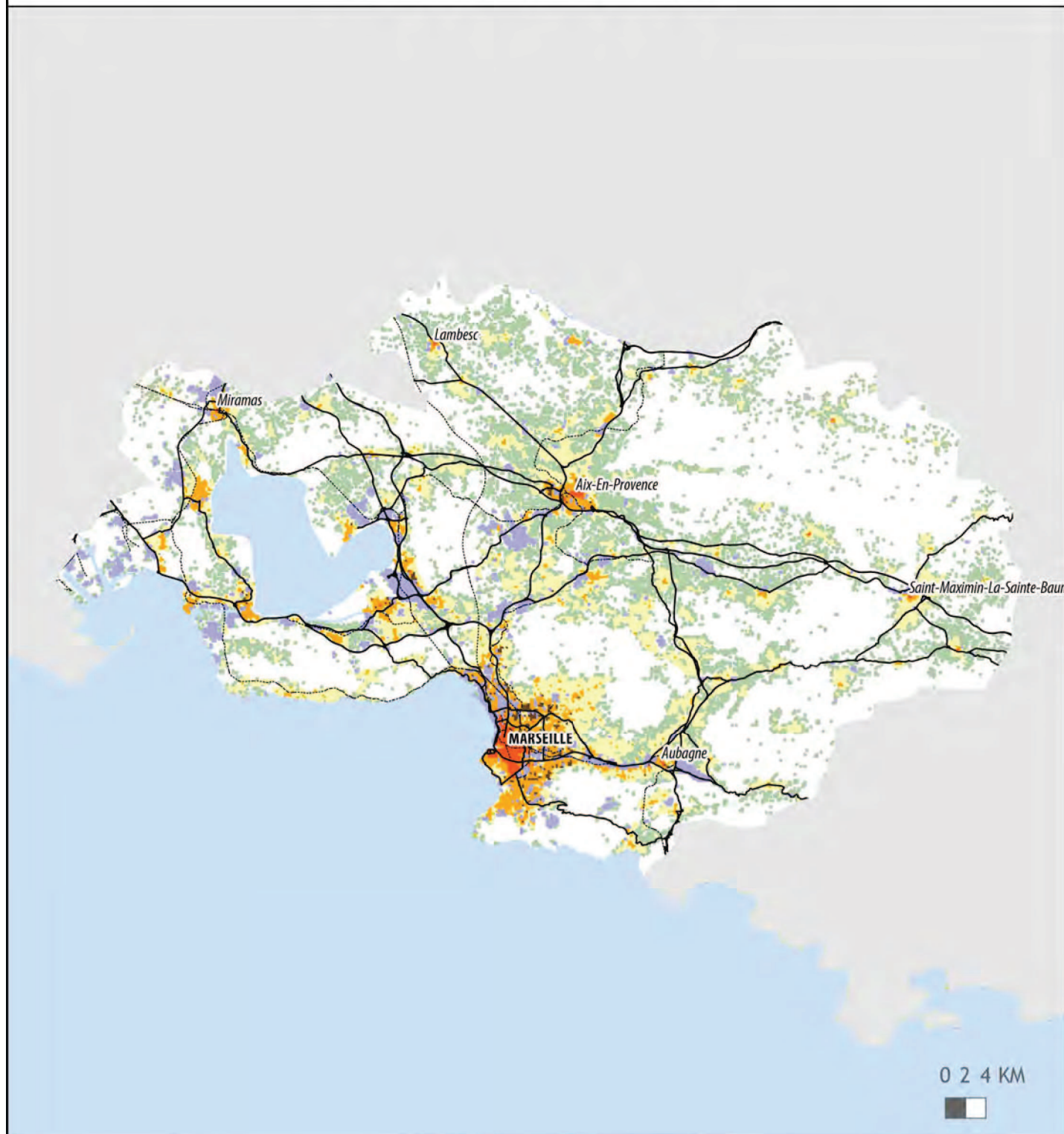


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE MARSEILLE - AIX-EN-PROVENCE



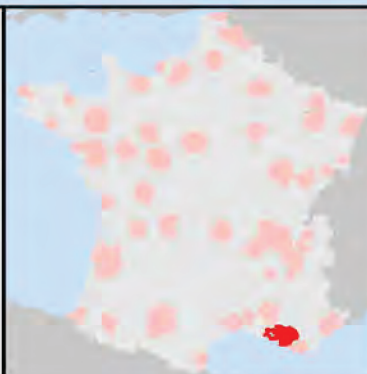
LEGENDE

TISSUS URBAINS

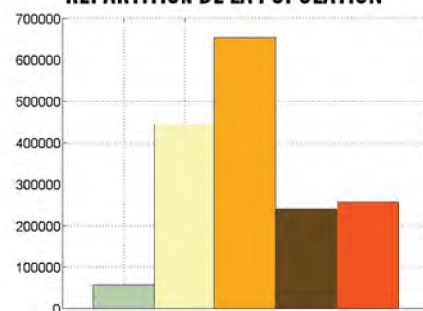
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau

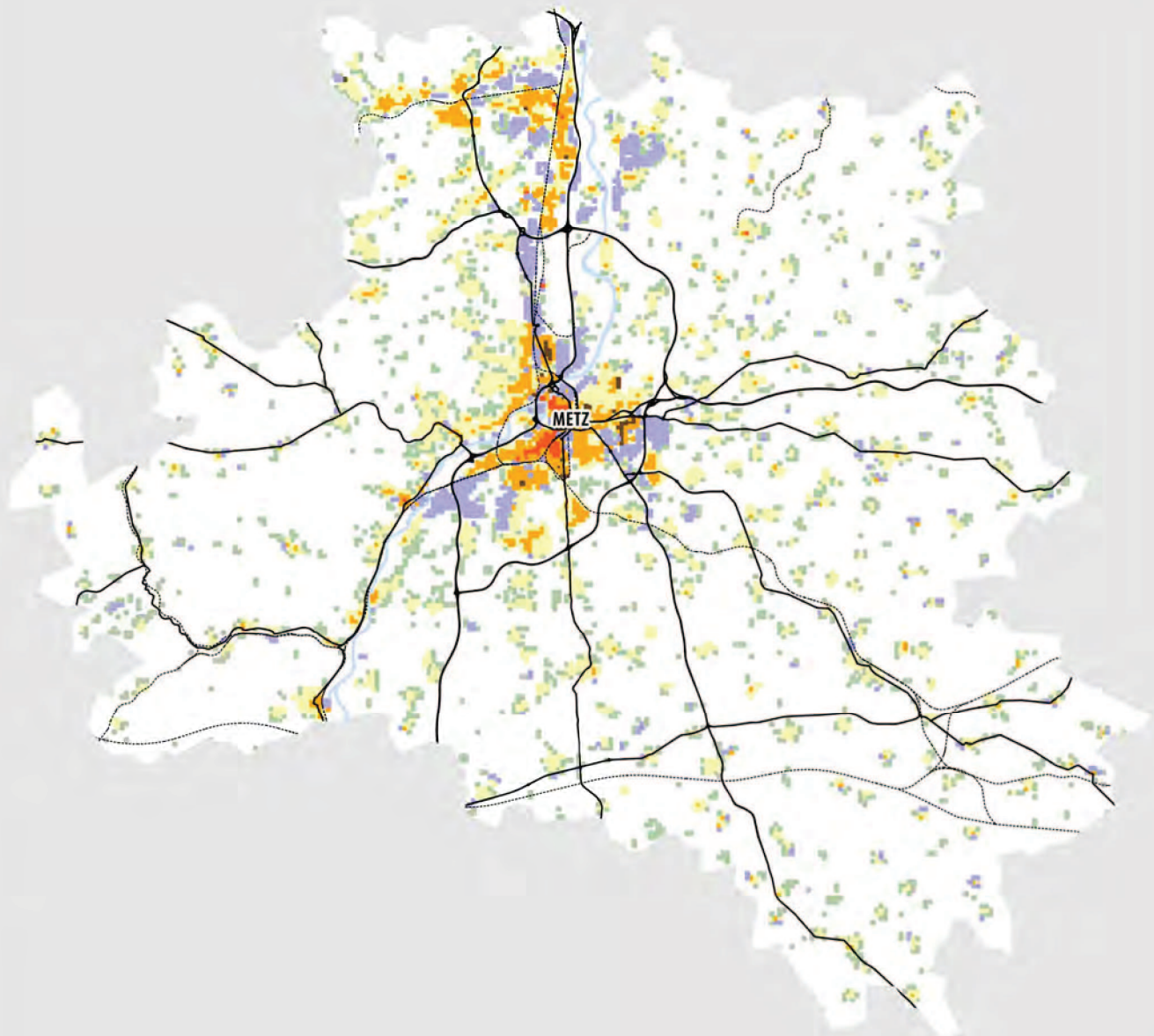


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE METZ



0 2 4 KM



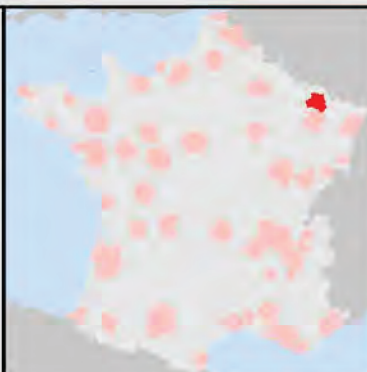
LEGENDE

TISSUS URBAINS

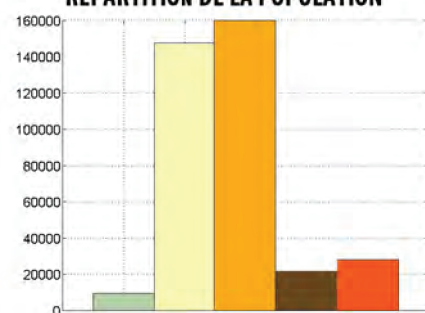
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau

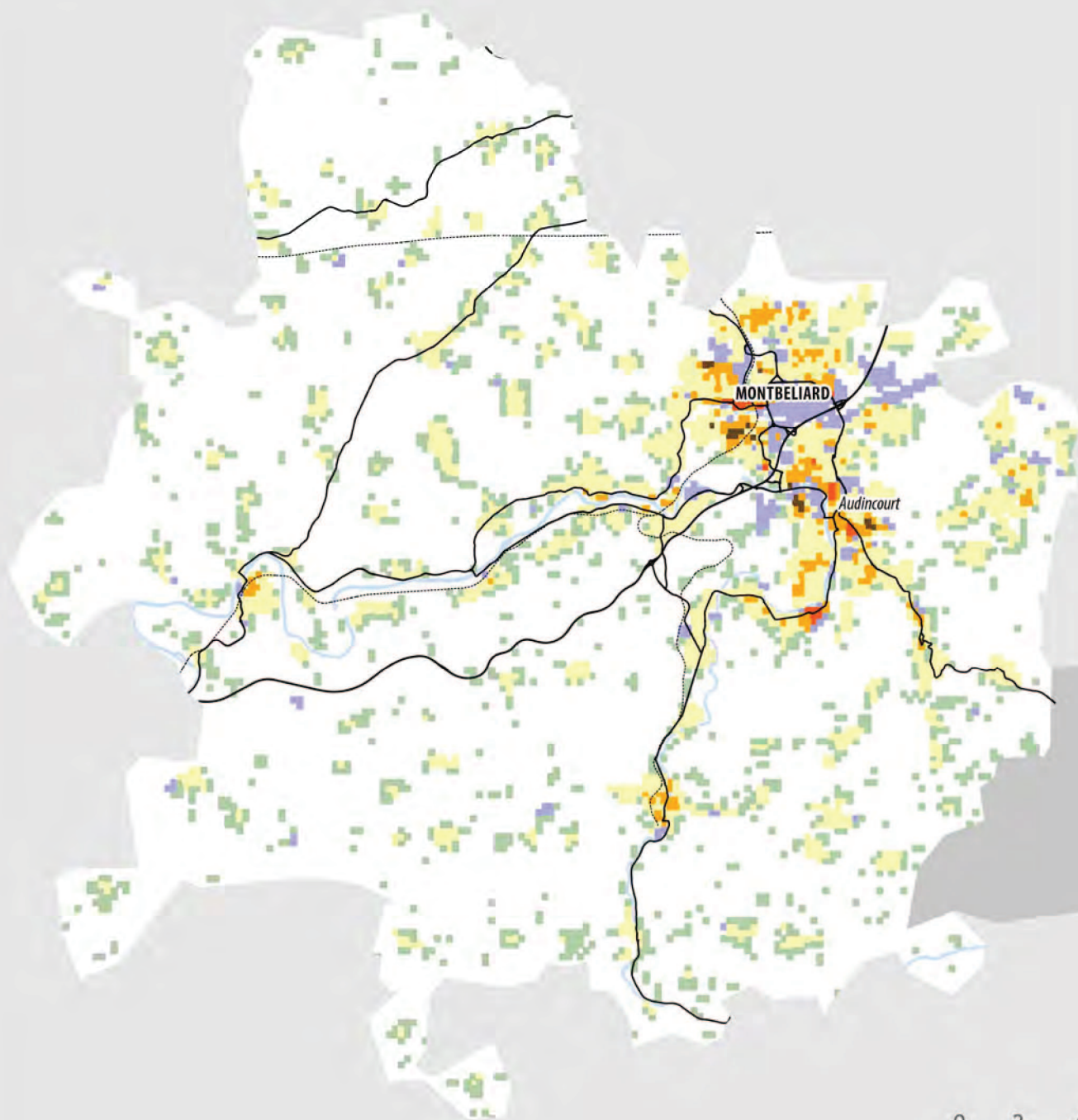


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE MONTBÉLIARD



0 2 4 KM

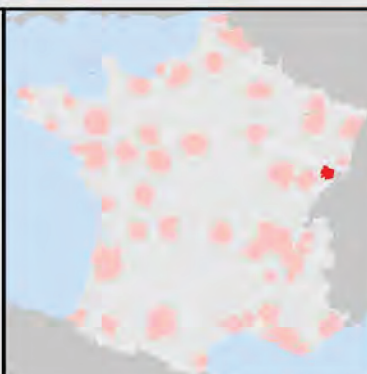
LEGENDE

TISSUS URBAINS

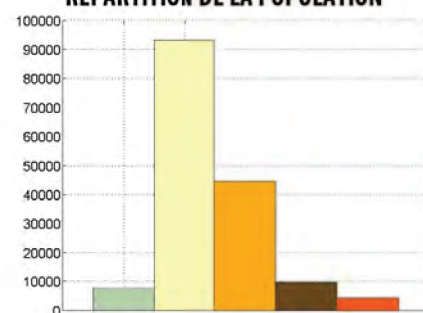
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- · — · — Voie ferrée
- Route principale
- Surface en eau

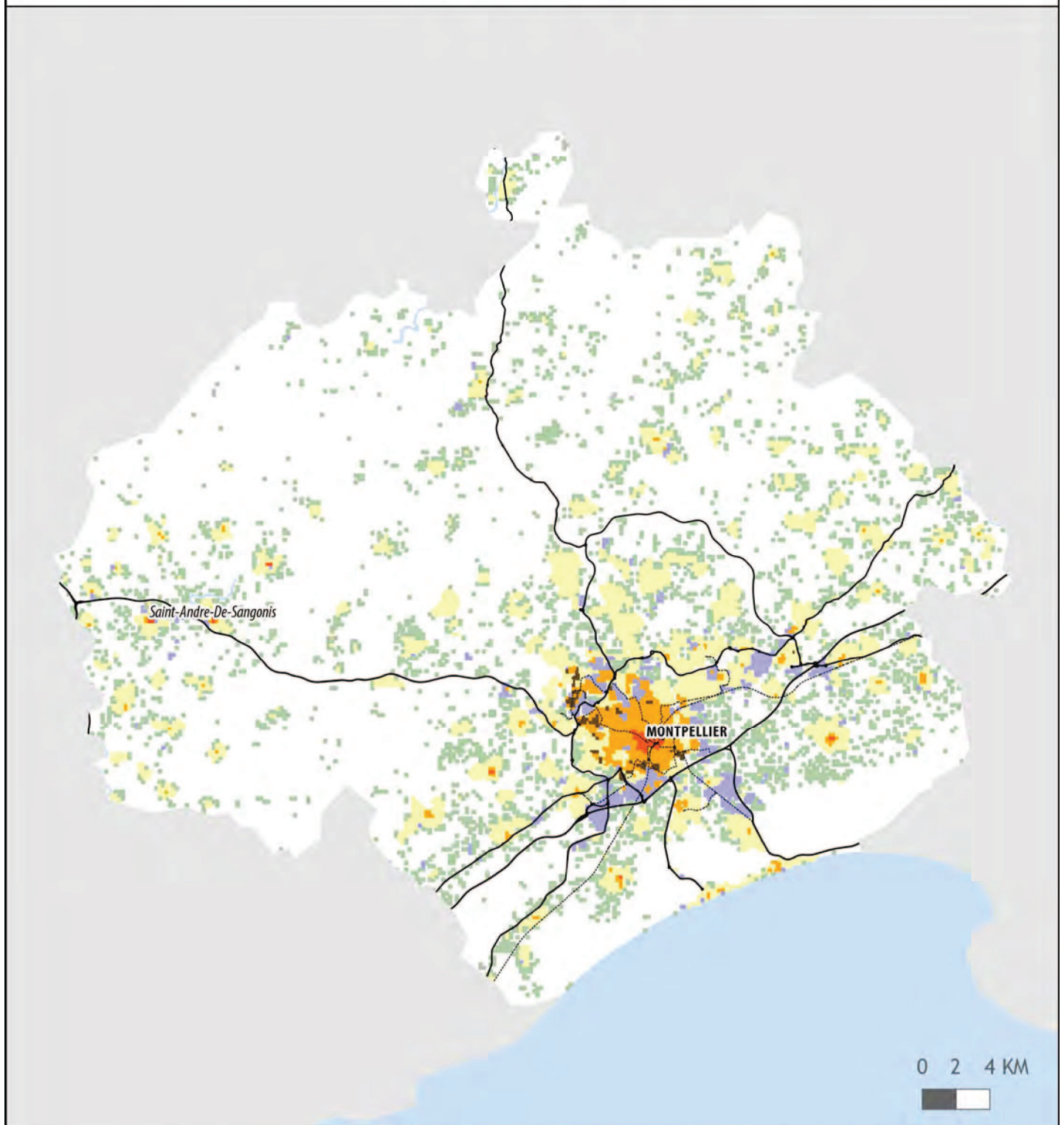


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE MONTPELLIER



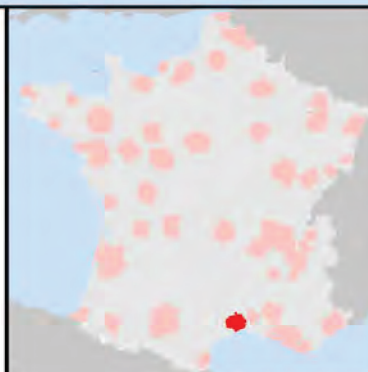
LEGENDE

TISSUS URBAINS

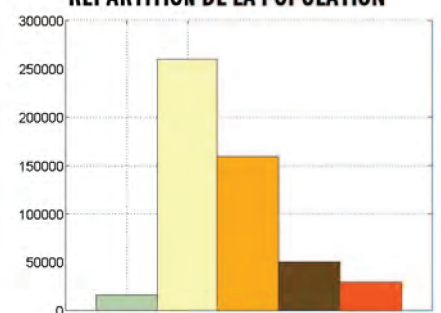
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

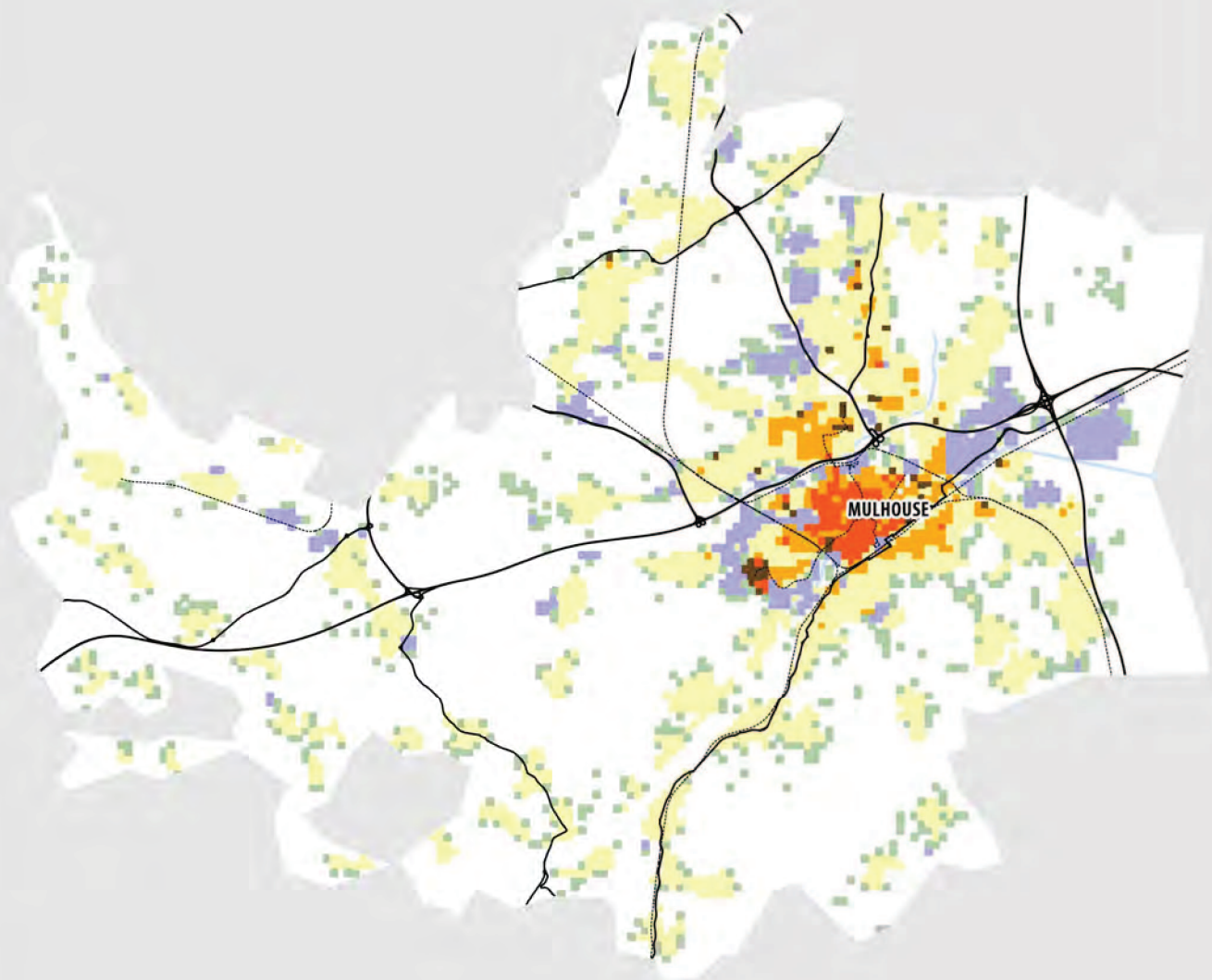


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE MULHOUSE



0 2 4 KM

LEGENDE

TISSUS URBAINS

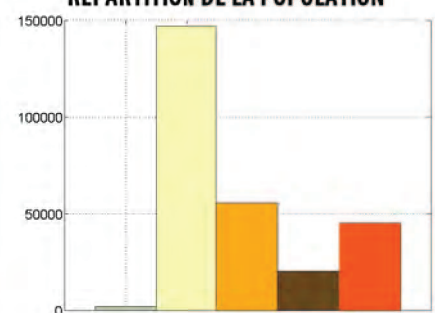
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE NANCY



0 2 4 KM

LEGENDE

TISSUS URBAINS

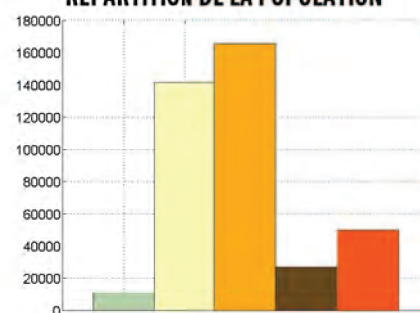
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau

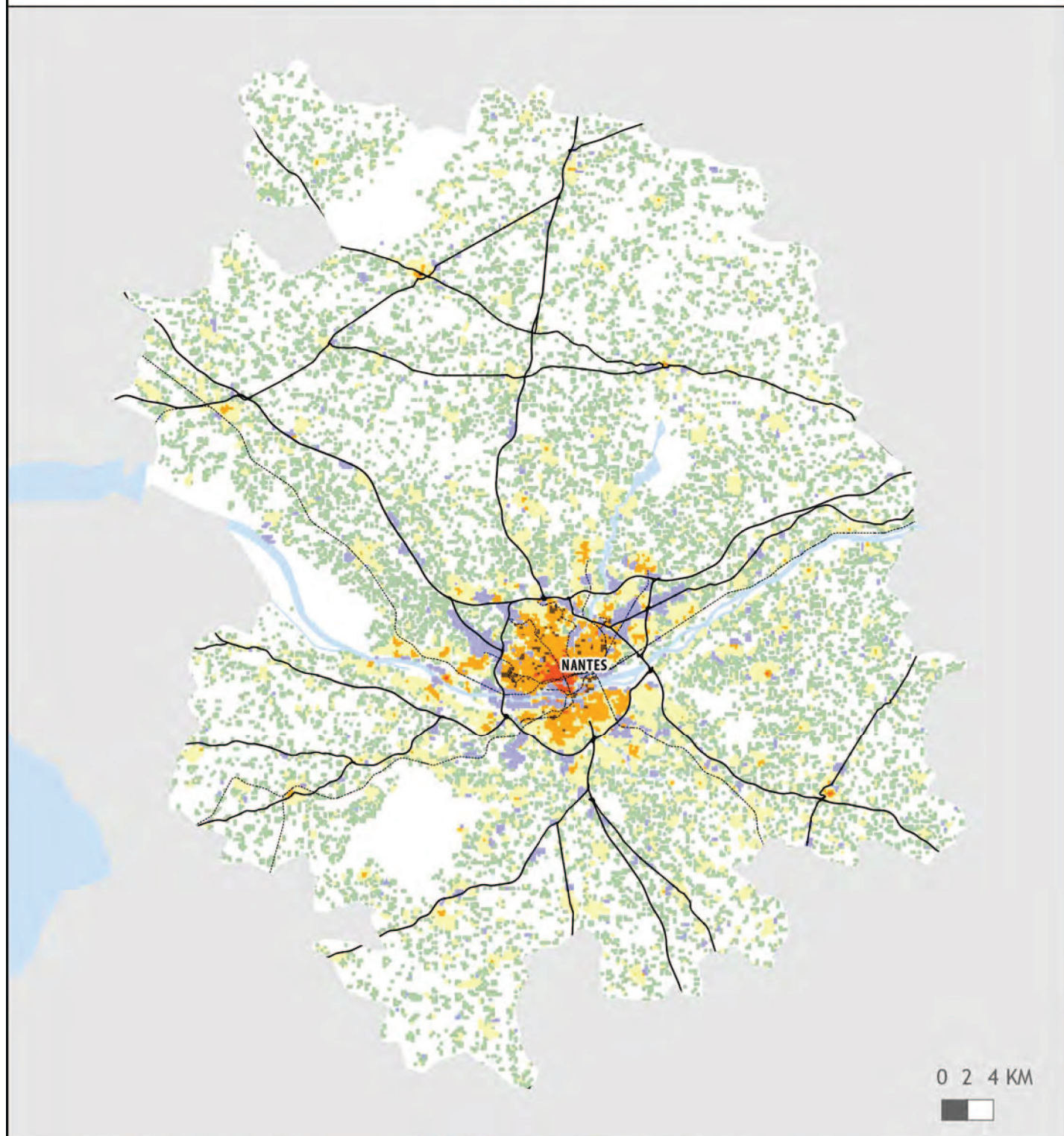


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE NANTES



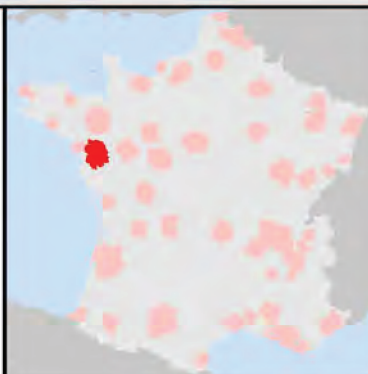
LEGENDE

TISSUS URBAINS

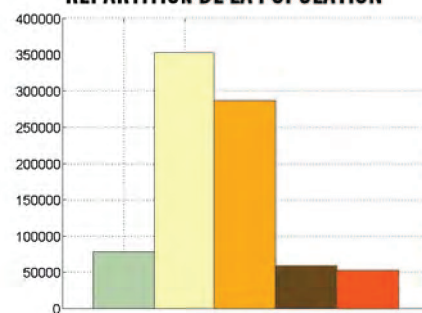
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau

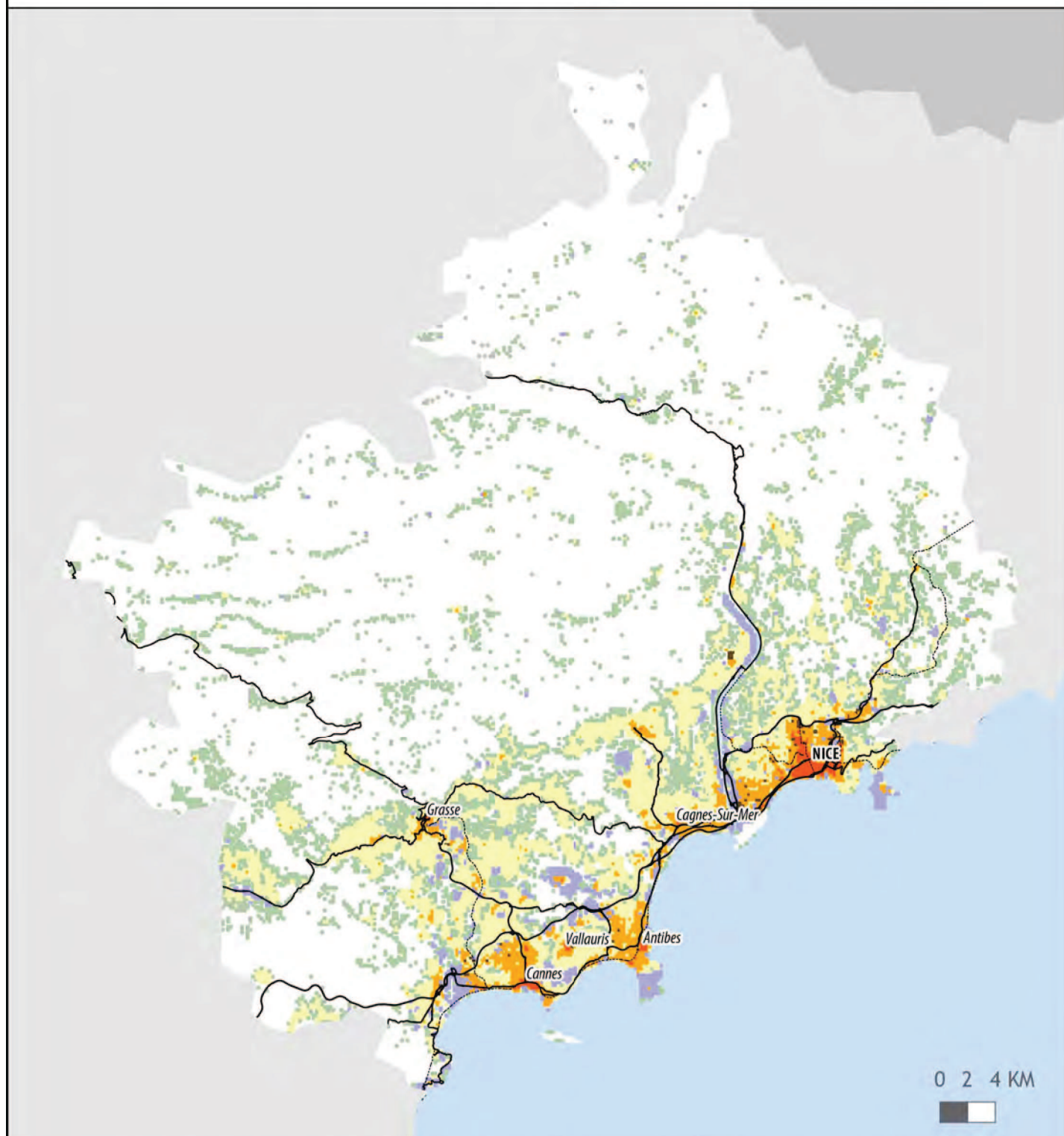


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE NICE



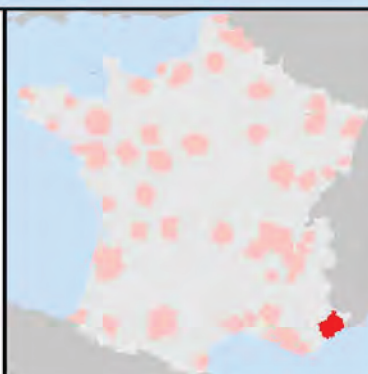
LEGENDE

TISSUS URBAINS

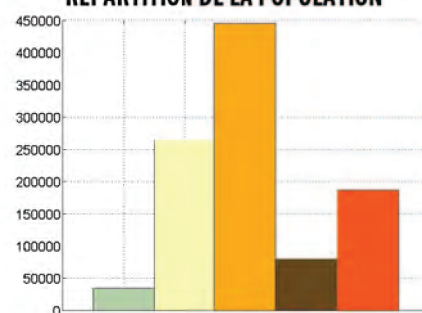
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE NÎMES



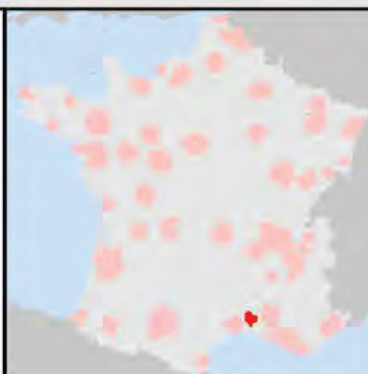
LEGENDE

TISSUS URBAINS

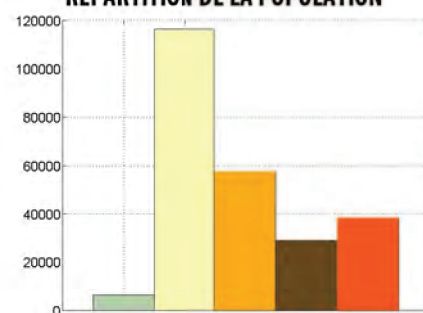
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE D'ORLÉANS



0 2 4 KM

LEGENDE

TISSUS URBAINS

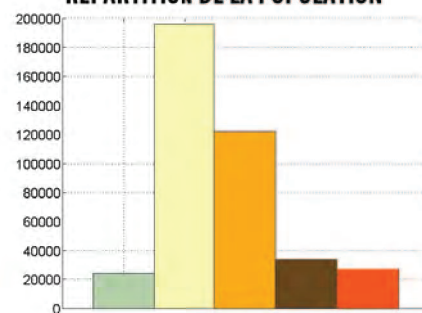
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE PAU



LEGENDE

TISSUS URBAINS

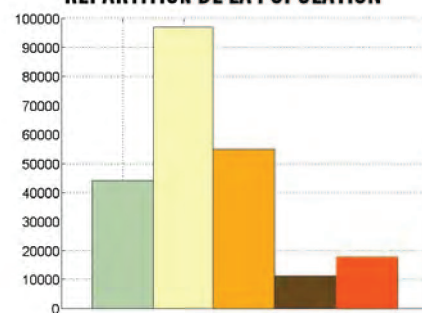
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau

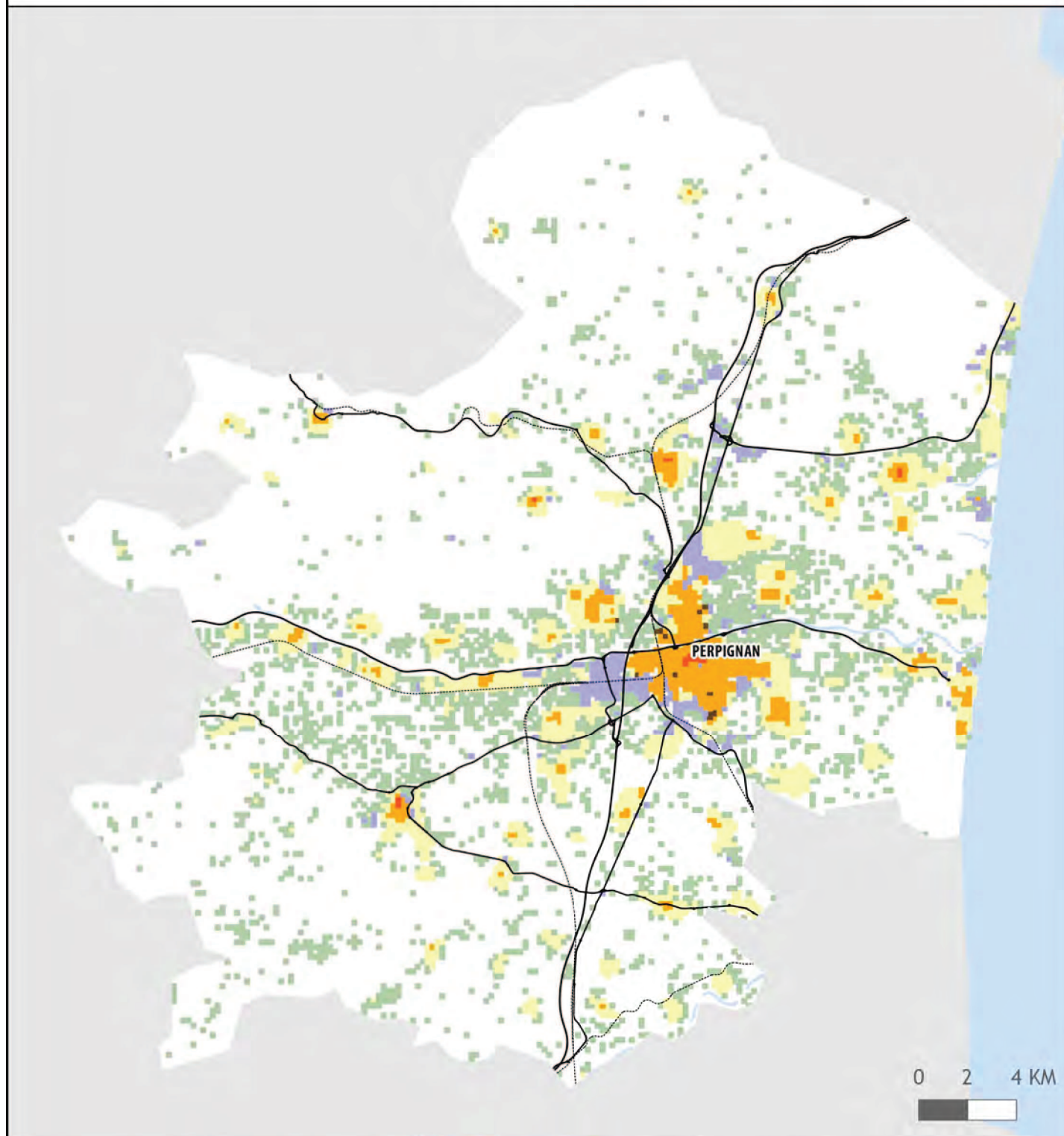


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE PERPIGNAN



LEGENDE

TISSUS URBAINS

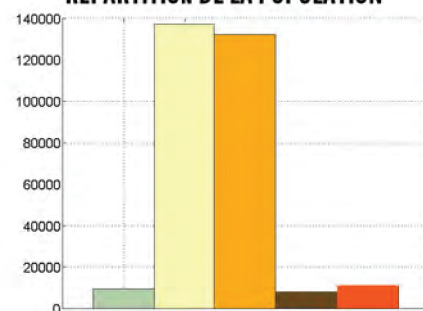
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE POITIERS



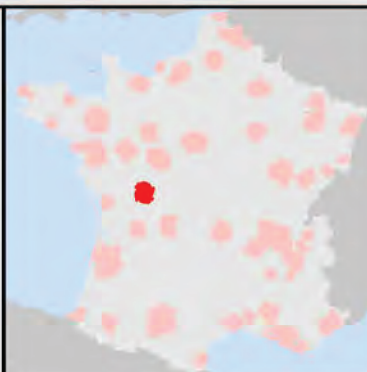
LEGENDE

TISSUS URBAINS

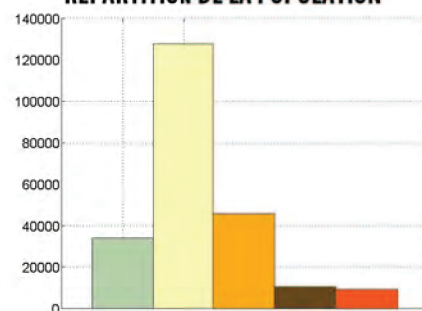
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE REIMS



LEGENDE

TISSUS URBAINS

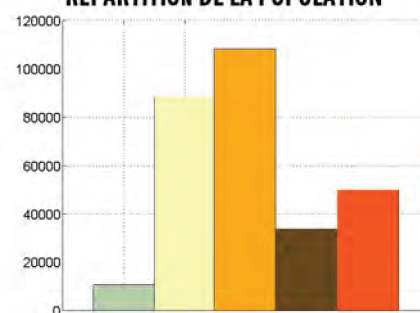
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE RENNES



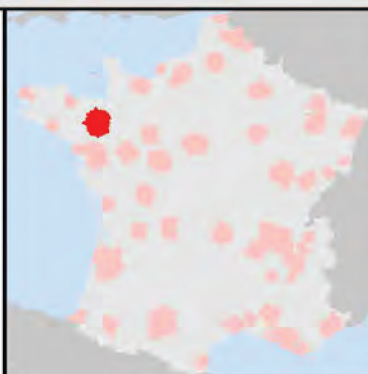
LEGENDE

TISSUS URBAINS

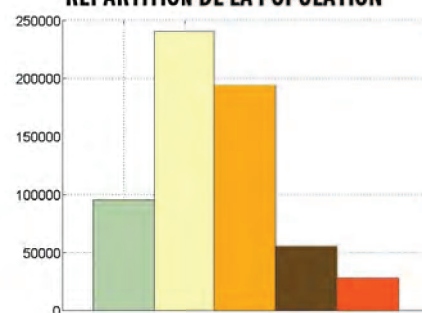
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE ROUEN



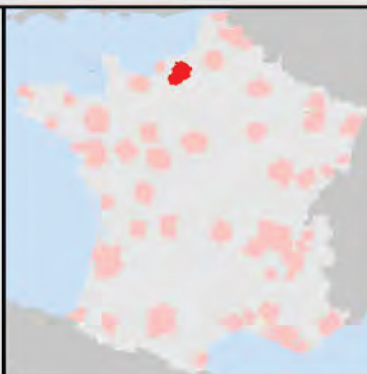
LEGENDE

TISSUS URBAINS

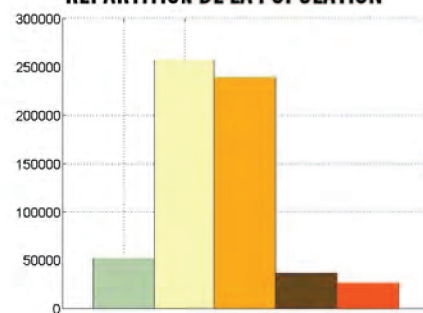
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- ⋯ Voie ferrée
- Route principale
- Surface en eau

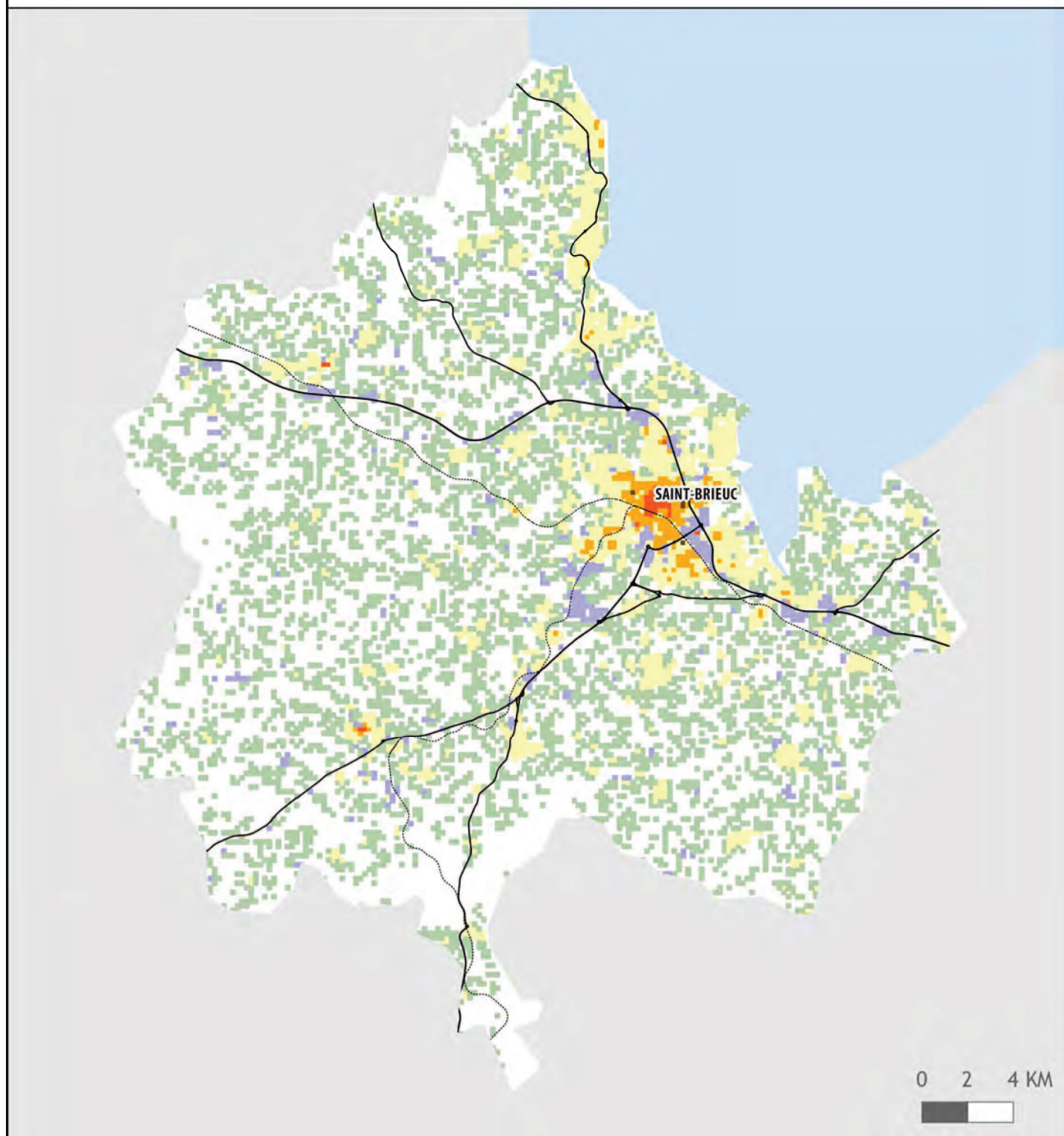


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE SAINT-BRIEUC



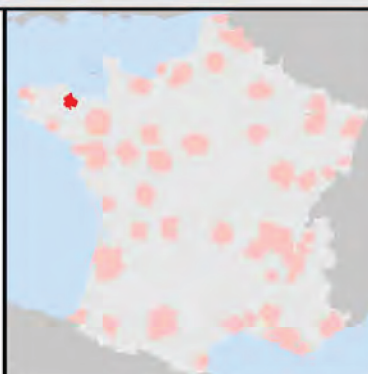
LEGENDE

TISSUS URBAINS

- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau

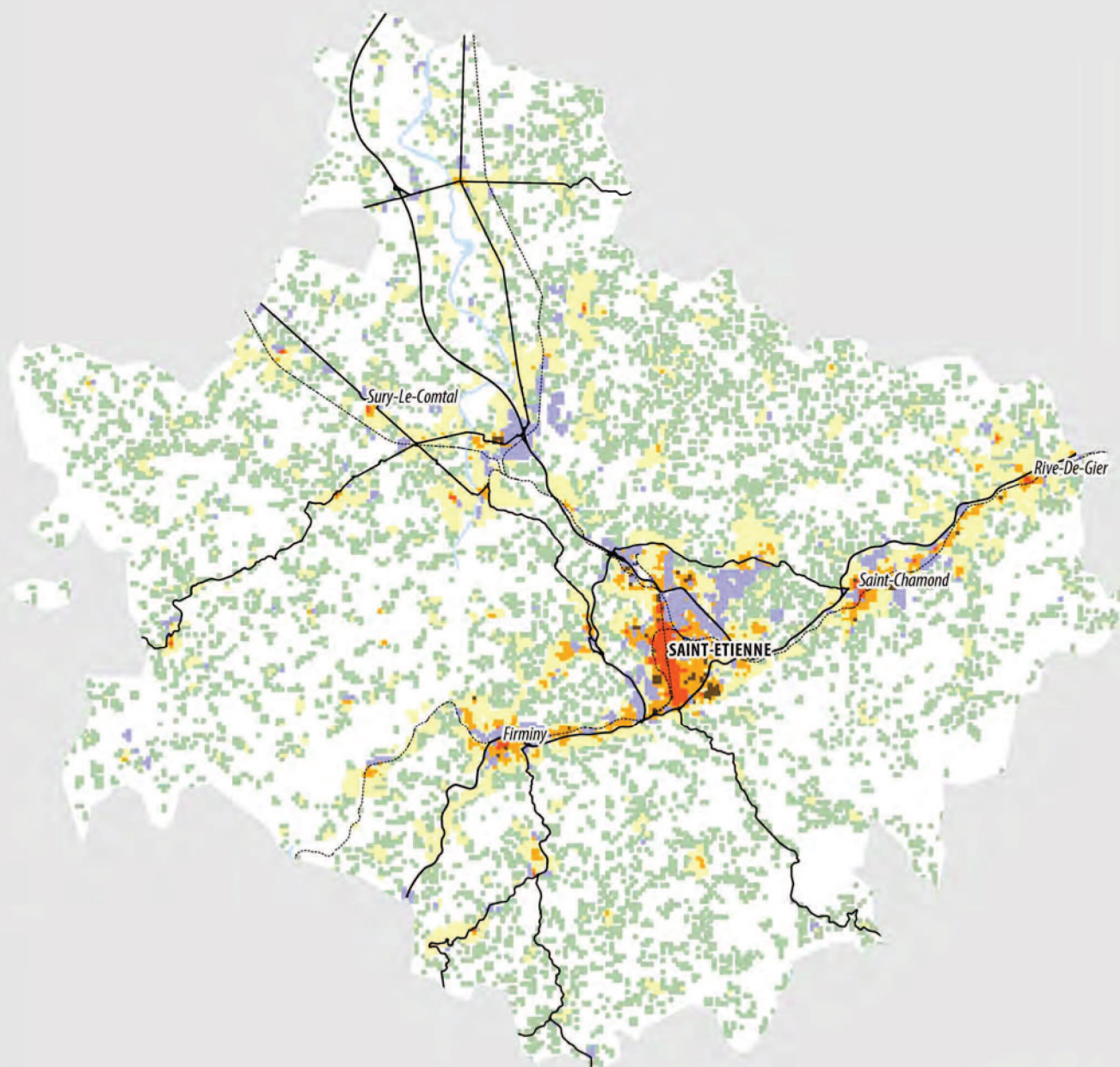


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE SAINT-ÉTIENNE



0 2 4 KM



LEGENDE

TISSUS URBAINS

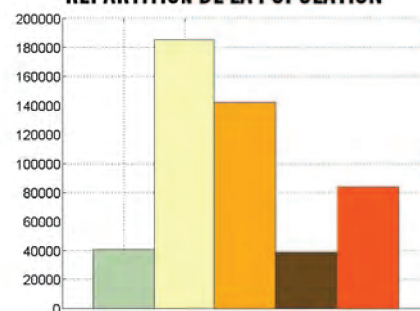
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

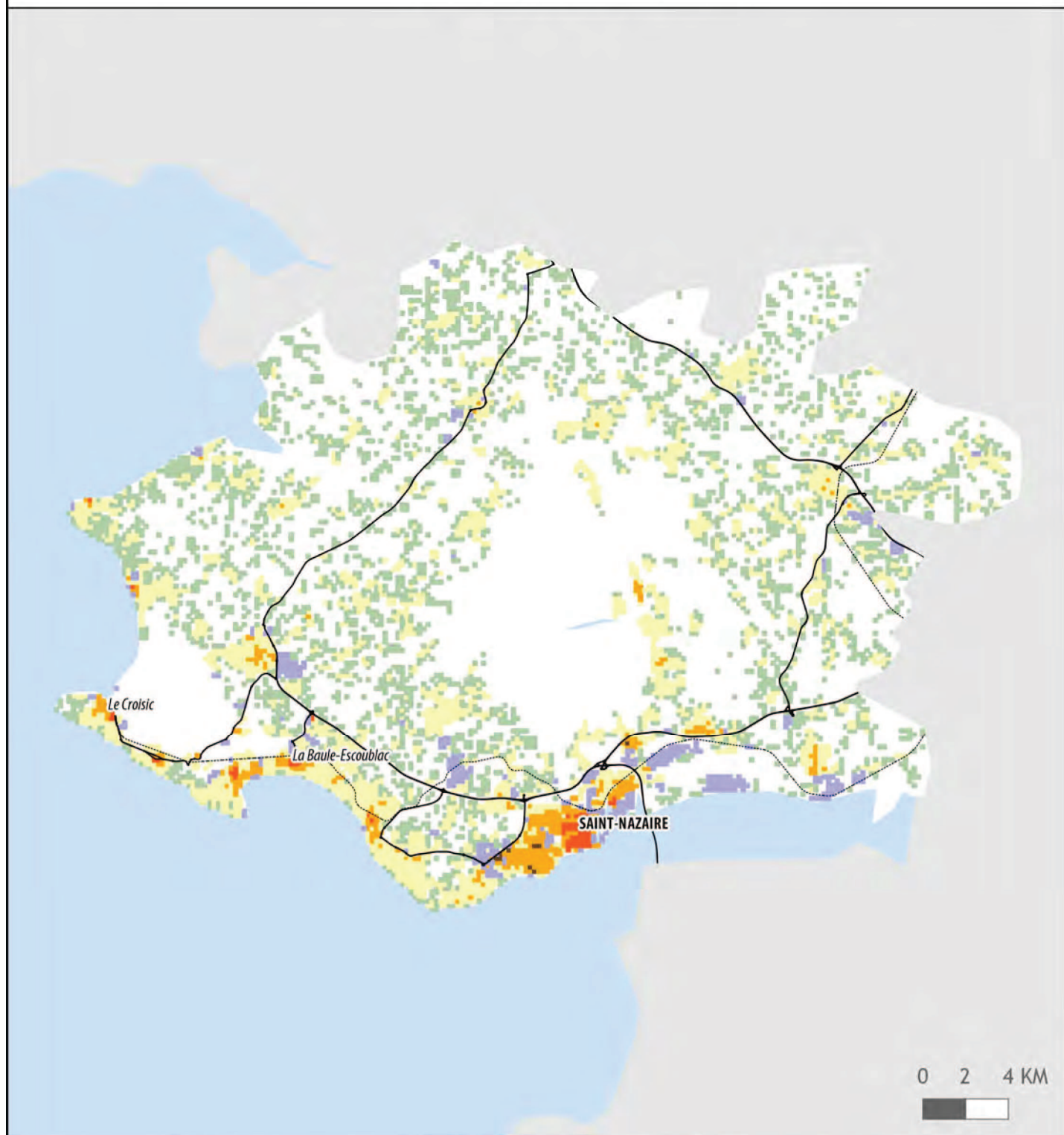


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE SAINT-NAZAIRE



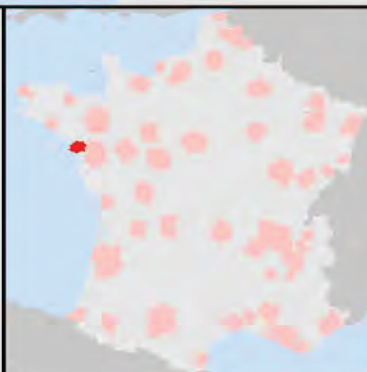
LEGENDE

TISSUS URBAINS

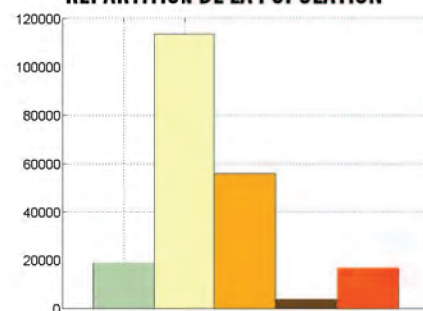
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- ⋯ Voie ferrée
- Route principale
- Surface en eau

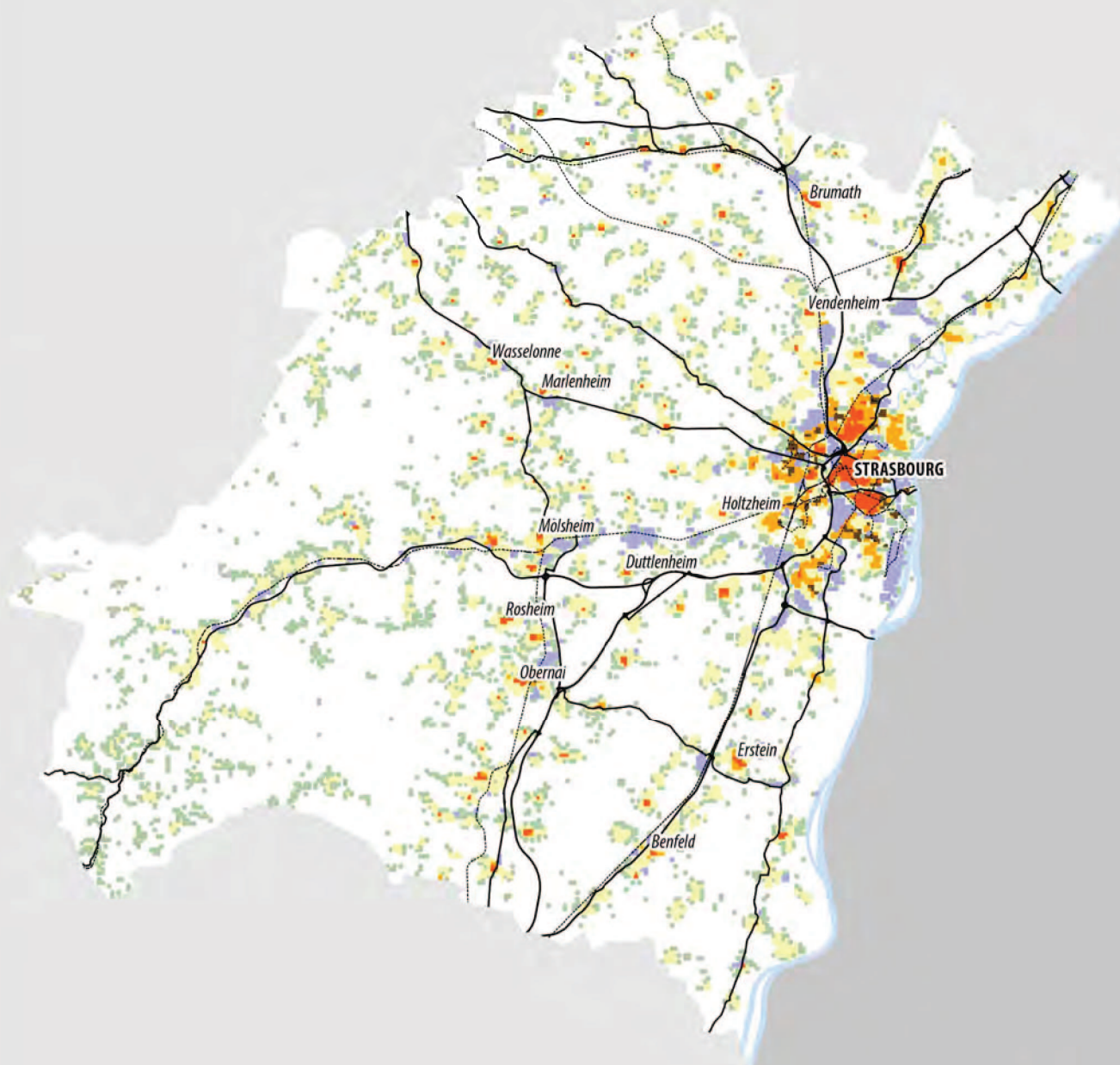


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE STRASBOURG (PARTIE FRANÇAISE)



0 2 4 KM



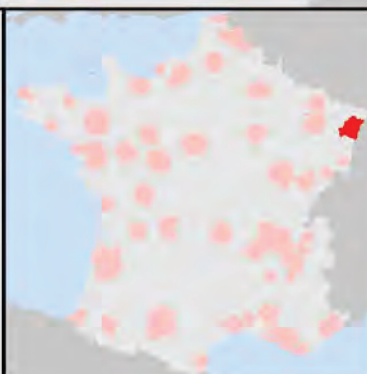
LEGENDE

TISSUS URBAINS

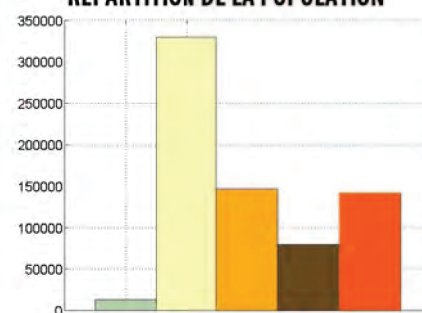
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

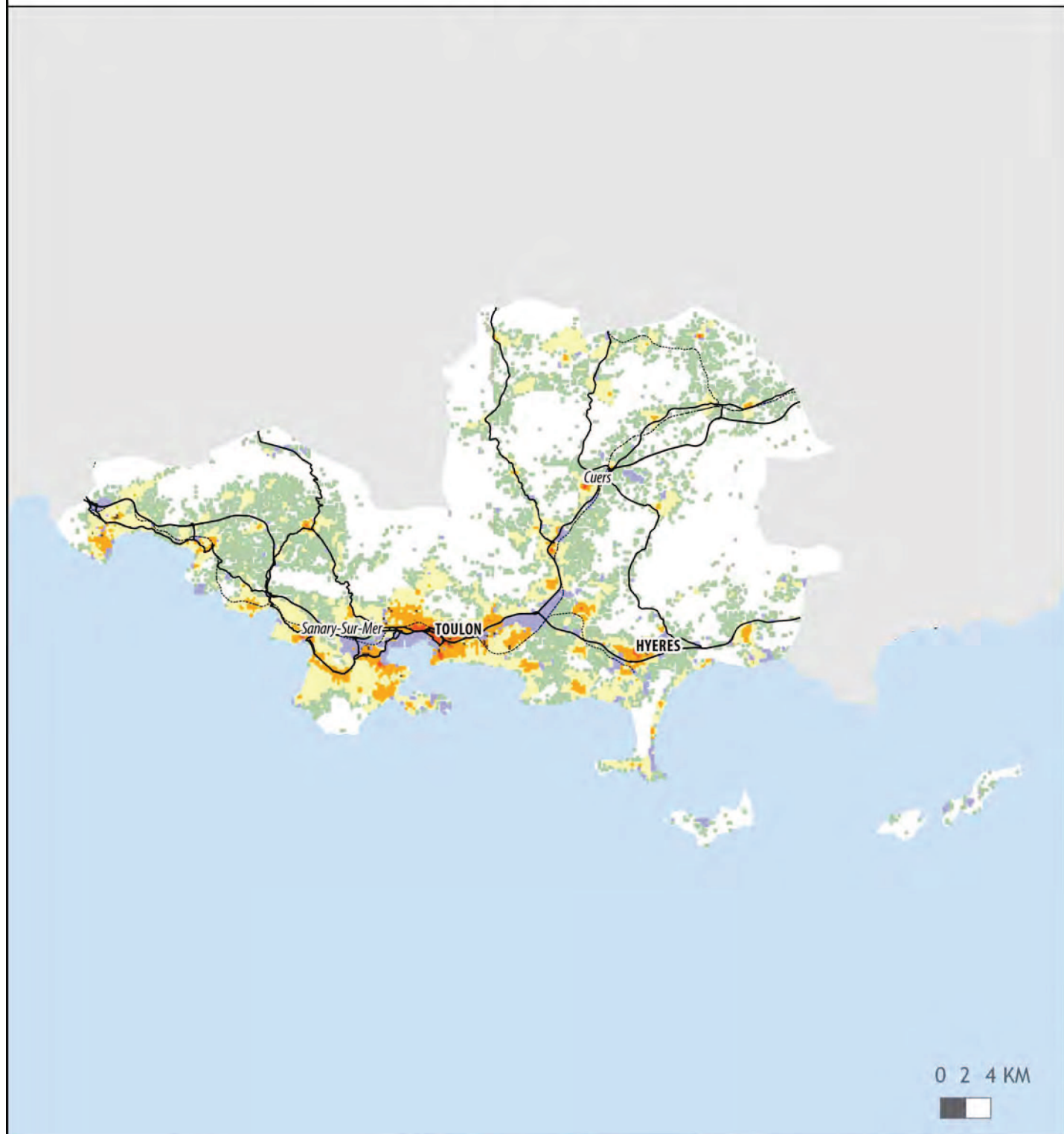


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE TOULON



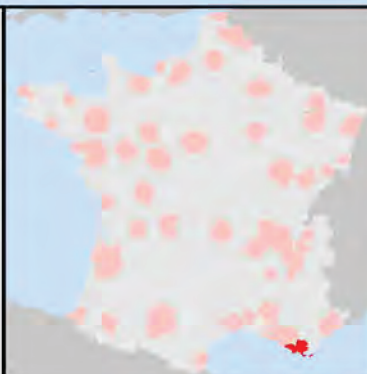
LEGENDE

TISSUS URBAINS

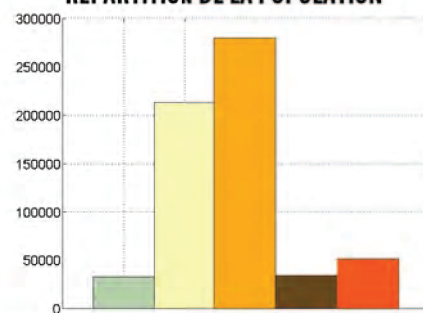
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- - - Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE TOULOUSE



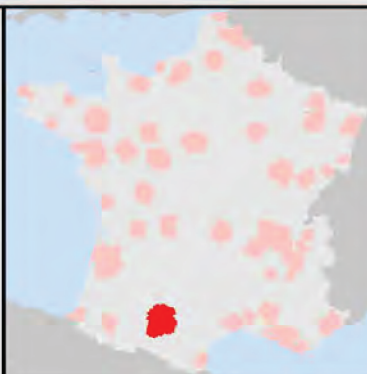
LEGENDE

TISSUS URBAINS

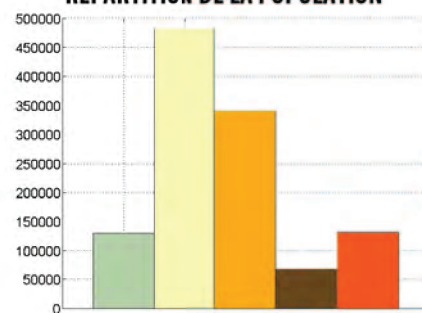
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- · — · — Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE TOURS



0 2 4 KM



LEGENDE

TISSUS URBAINS

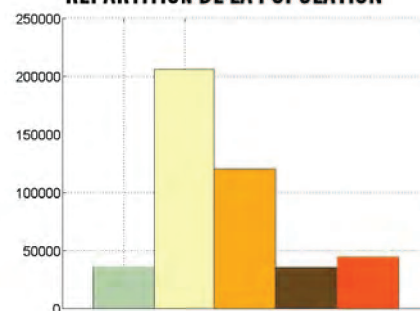
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE TROYES



0 2 4 KM



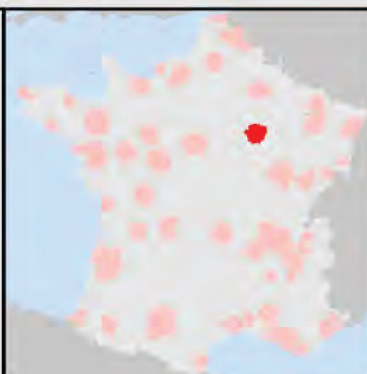
LEGENDE

TISSUS URBAINS

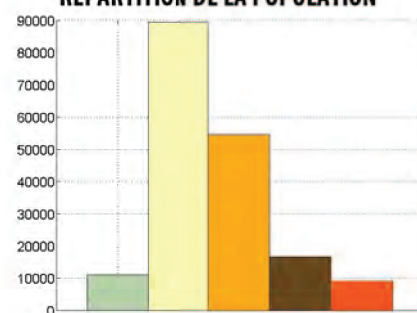
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

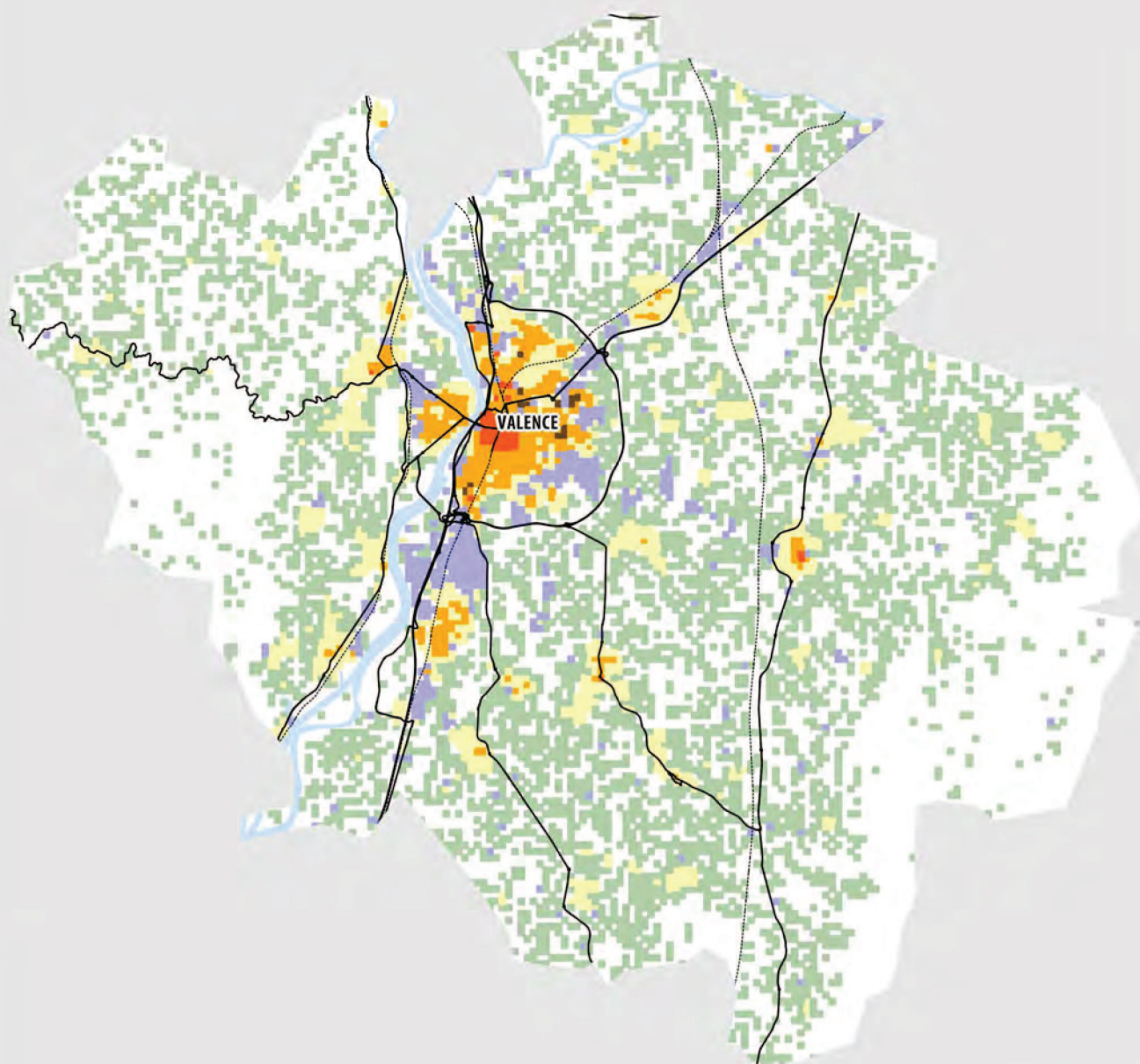


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE VALENCE



0 2 4 KM



LEGENDE

TISSUS URBAINS

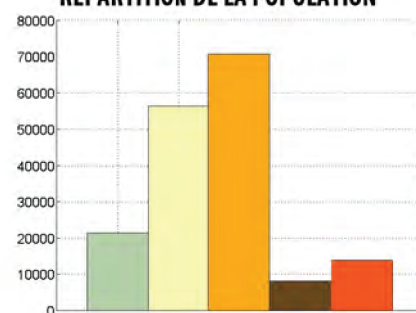
- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

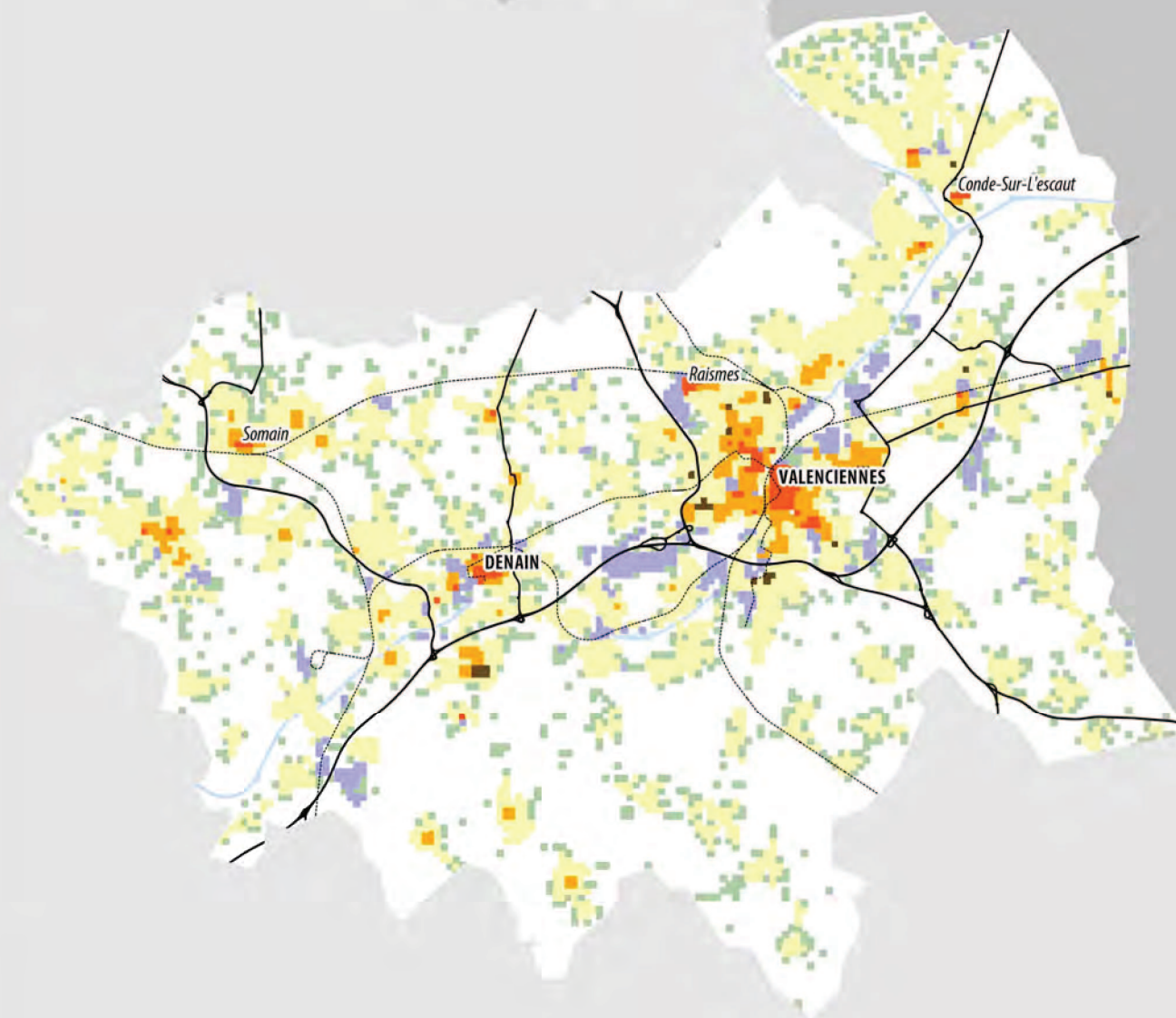


REPARTITION DE LA POPULATION



ZONAGE DES TISSUS URBAINS EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE VALENCIENNES (PARTIE FRANÇAISE)



0 2 4 KM

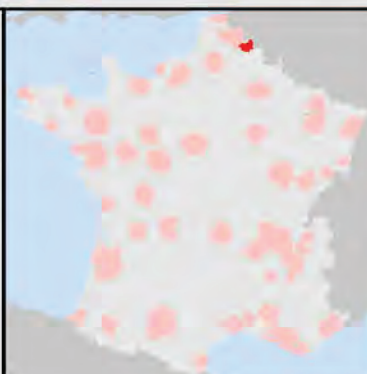
LEGENDE

TISSUS URBAINS

- Centre historique
- Habitat collectif
- Habitat mixte et dense
- Habitat individuel groupé
- Habitat individuel dispersé
- Zone d'activité

HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION

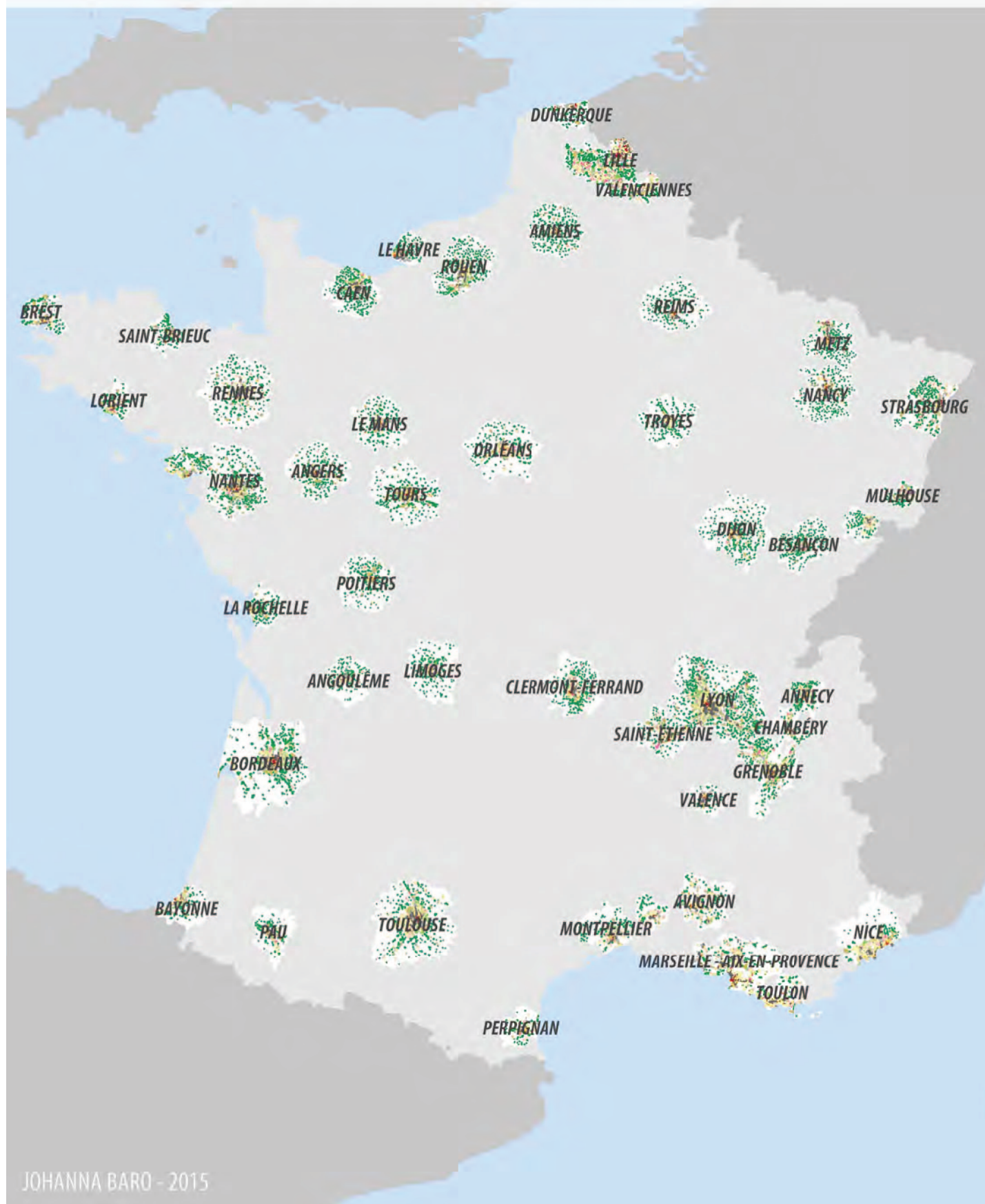


B.2 Atlas des morphotypes

Référence permanent de l'atlas : BARO, Johanna, 2015, « Atlas des morphotypes de 50 aires urbaines françaises », <http://hdl.handle.net/TEST/10028> Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux [Distributor]

ATLAS DES MORPHOTYPES

DE 50 AIRES URBAINES FRANÇAISES



Atlas des morphotypes de 50 aires urbaines françaises

Cet atlas propose une cartographie de la morphologie des 50 plus grandes aires urbaines françaises, exception faite de Paris. Il fournit une vision de la morphologie urbaine selon une nomenclature détaillée de quatorze morphotypes décrivant les logiques d'organisation interne des aires urbaines.

Ces catégories de morphotypes ont été identifiées à l'aide d'une modélisation chorématique dont l'objectif était de représenter différents processus caractérisant des structures urbaines comme les logiques de centre-périphérie, de sectorisation ou encore de hiérarchisation des lieux.

L'atlas révèle ainsi la diversité des configurations morphologiques en tenant compte de caractéristiques de forme, de position et de relation entre entités spatiales, qui permettent d'analyser la composition des aires urbaines en se référant à ce modèle chorématique d'organisation urbaine.

Les cartes présentées ci-après correspondent aux travaux de classifications tirés de la thèse en Sciences et Technologies de l'Information Géographique de Johanna Baro, intitulée « Modélisation multi-échelles de la morphologie urbaine à partir de données carroyées de population et de bâti ».

Ces classifications constituent un enrichissement de premières classifications en tissus urbains réalisées à partir de variables relatives à la population et aux surfaces bâties sur des grilles à mailles de 200 mètres de côté. Des règles de raisonnement spatial qualitatif ont été appliquées de manière systématique aux 50 aires urbaines étudiées pour traduire la spatialisation de ces morphotypes à partir de données.

Source d'habillage des cartes : BD Topo®, IGN

Auteur de l'atlas : Johanna Baro, johanna.baro@gmail.com

Mai 2015

STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE D'AMIENS

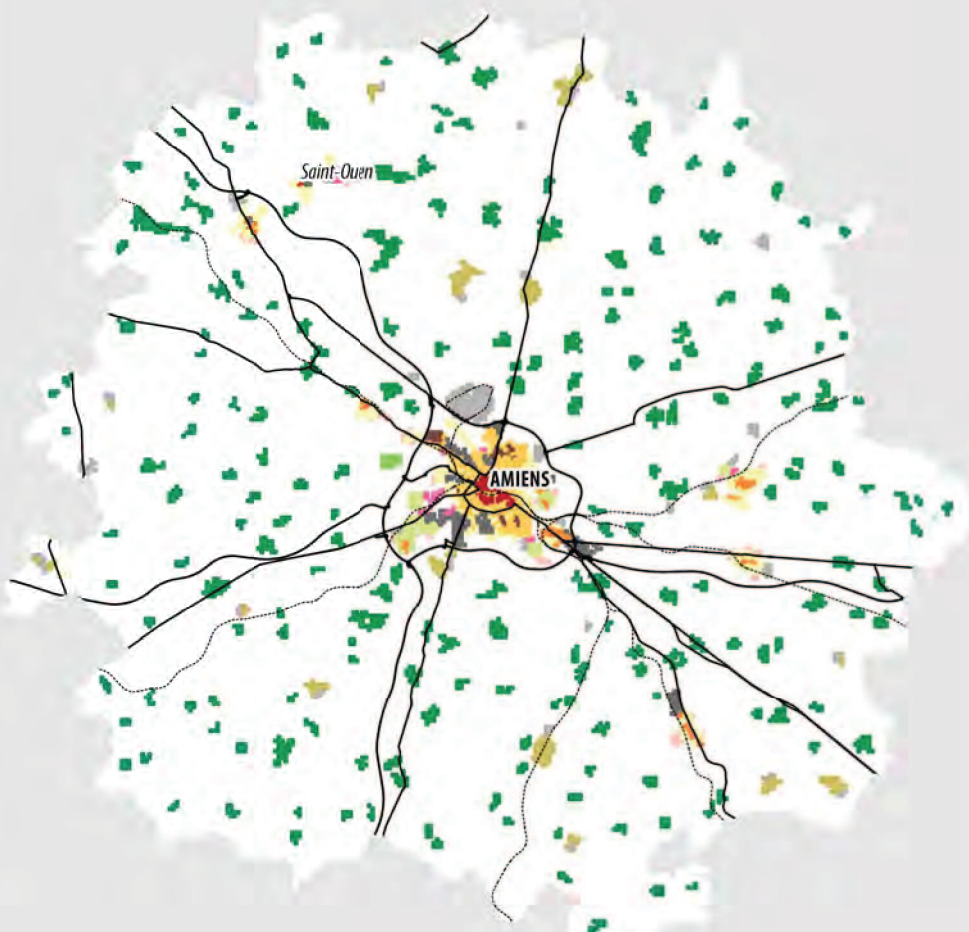
LEGENDE

MORPHOTYPES

- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

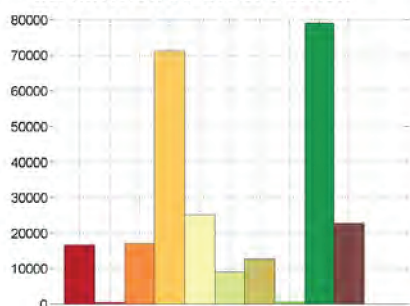
- Voie ferrée
- Route principale
- Surface en eau



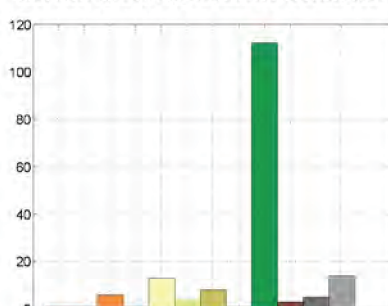
0 2 4 KM



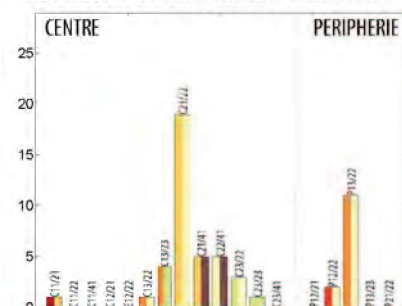
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE D'ANGERS

LEGENDE

MORPHOTYPES

- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

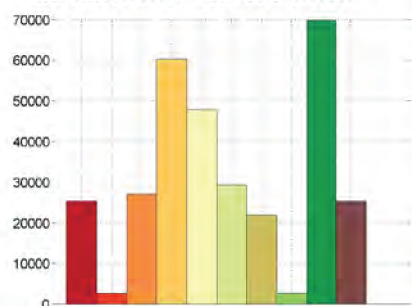
- Voie ferrée
- Route principale
- Surface en eau



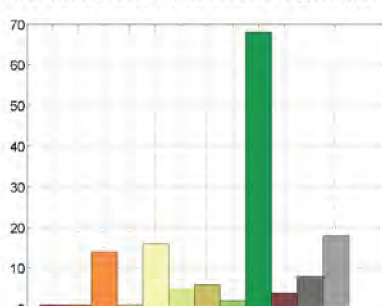
0 2 4 KM



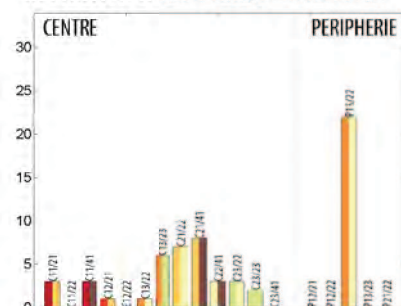
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE D'ANGOULÊME

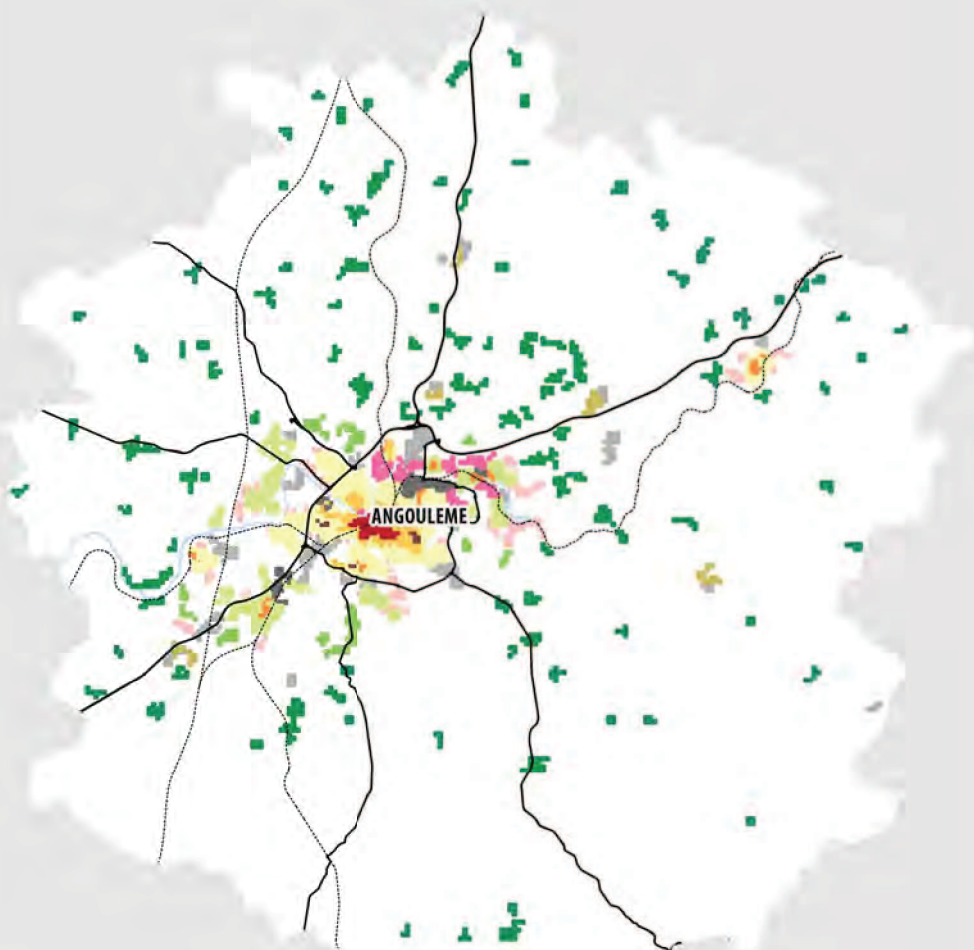
LEGENDE

MORPHOTYPES

- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

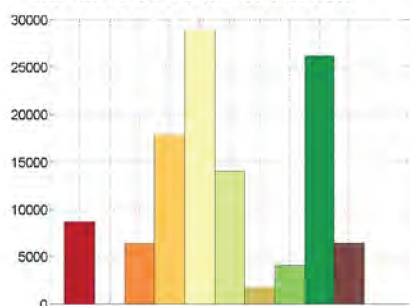
- Voie ferrée
- Route principale
- Surface en eau



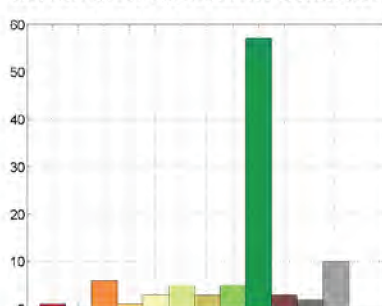
0 2 4 KM



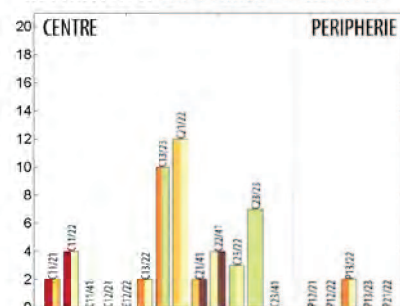
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE D'ANNECY

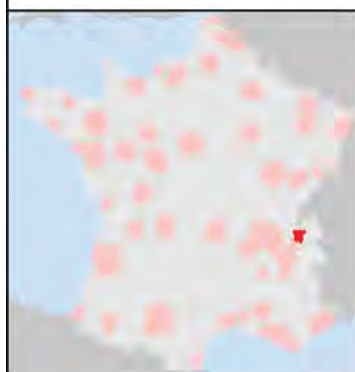
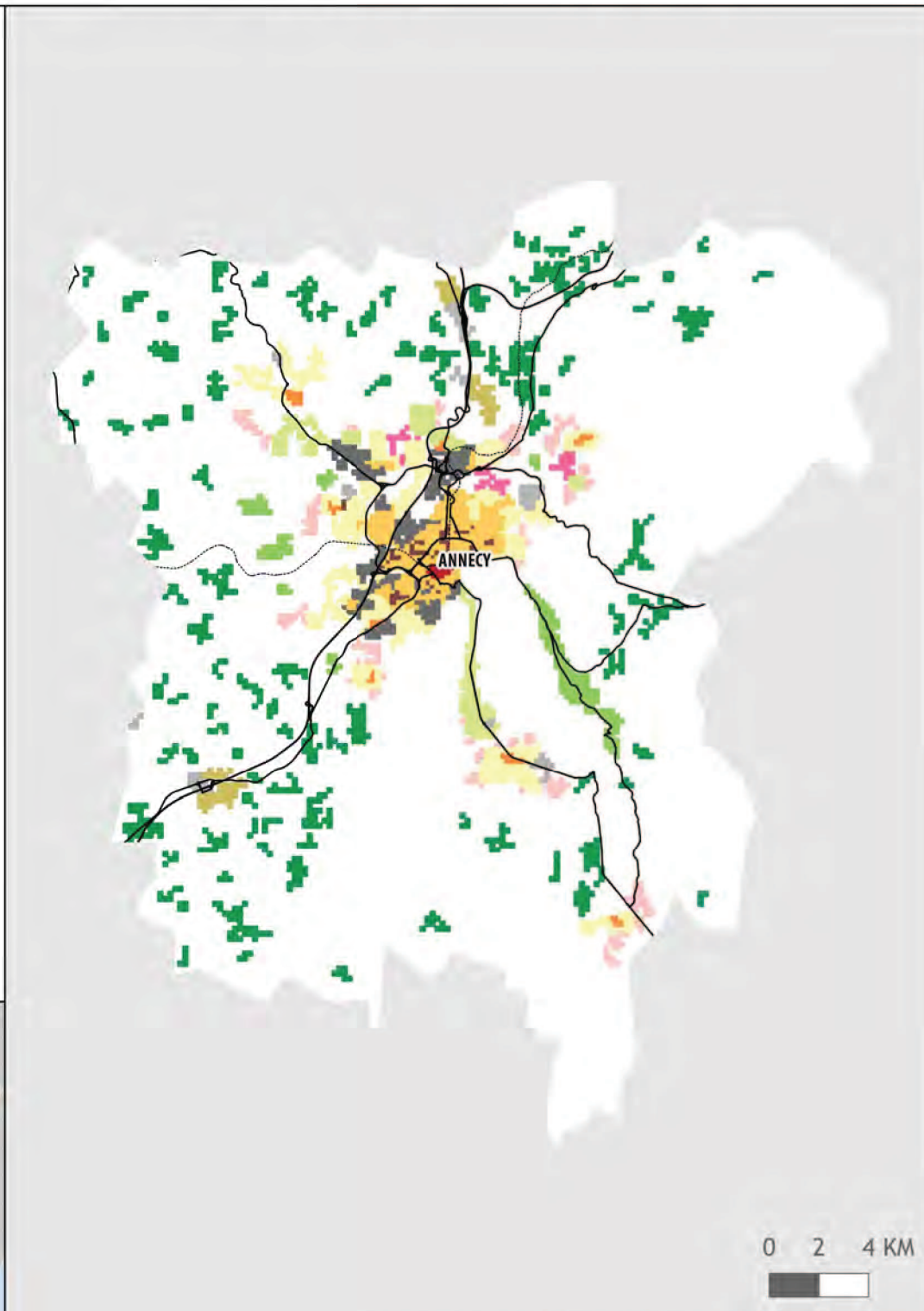
LEGENDE

MORPHOTYPES

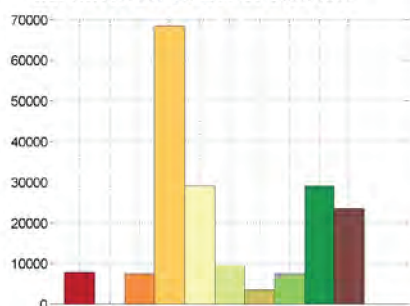
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

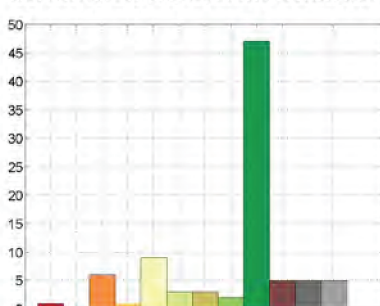
- Voie ferrée
- Route principale
- Surface en eau



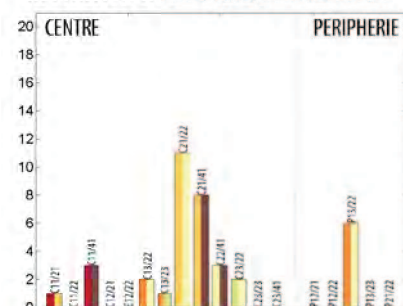
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE D'AVIGNON

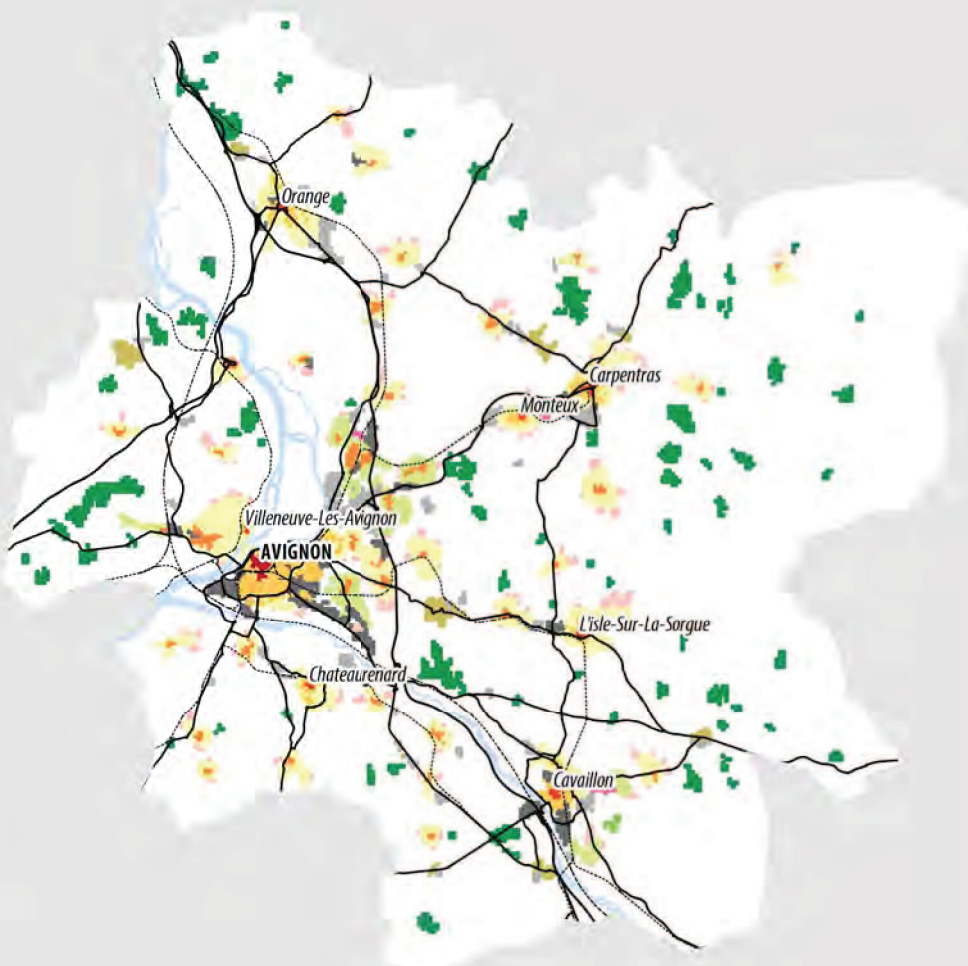
LEGENDE

MORPHOTYPES

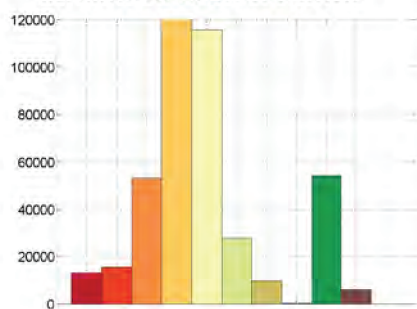
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

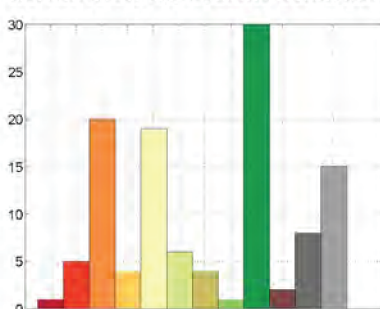
- Voie ferrée
- Route principale
- Surface en eau



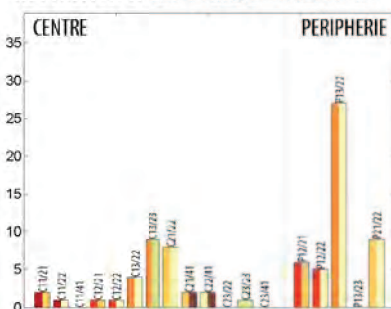
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE BAYONNE (PARTIE FRANÇAISE)

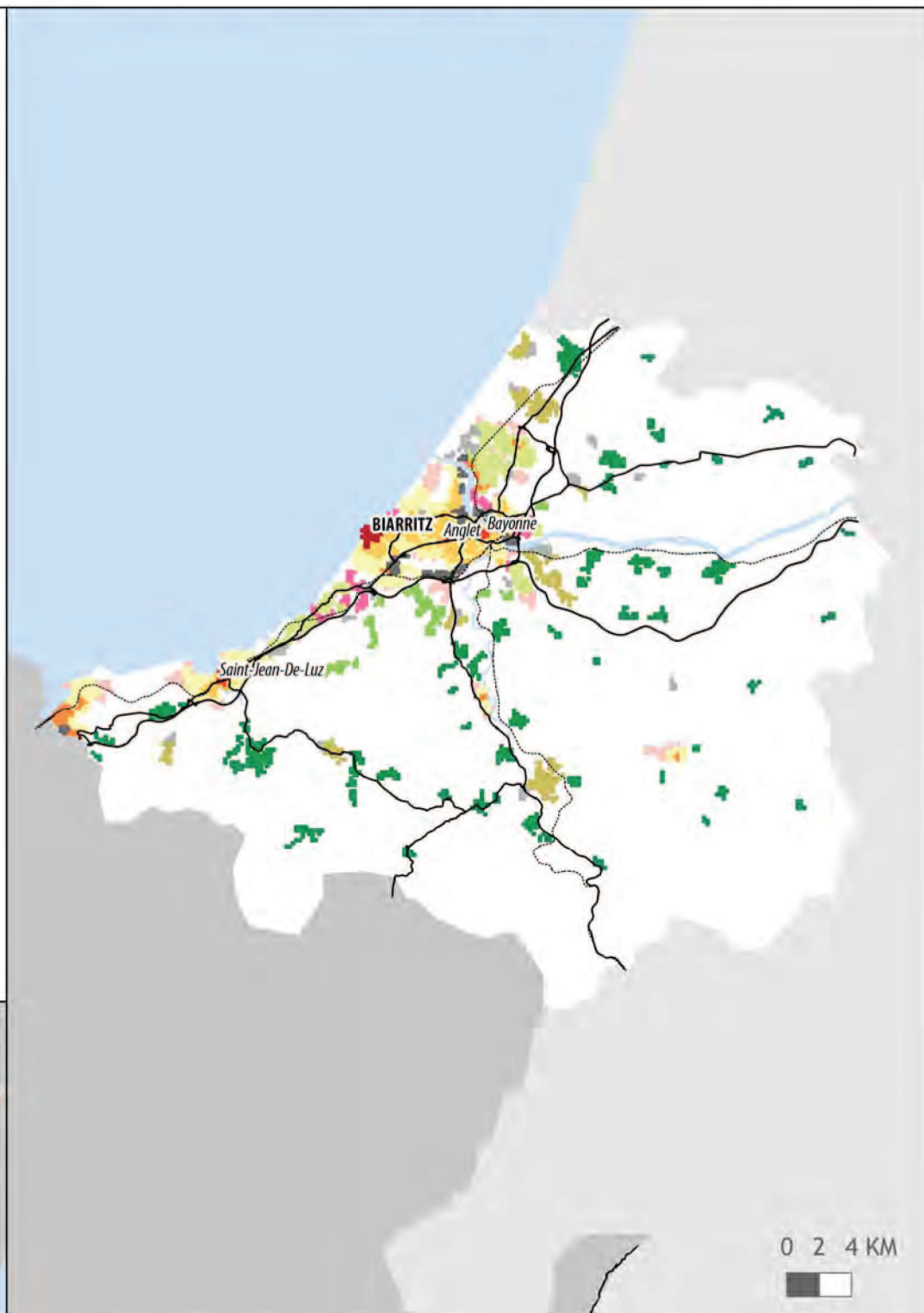
LEGENDE

MORPHOTYPES

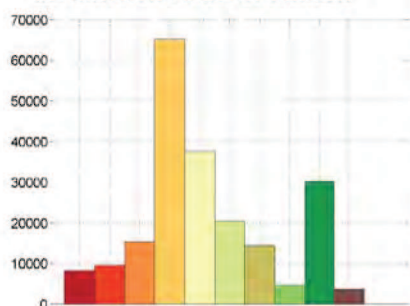
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

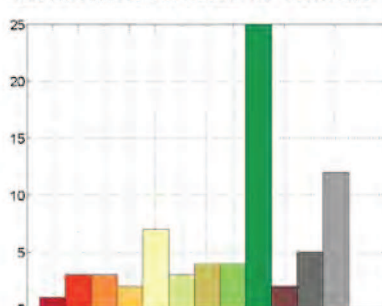
- Voie ferrée
- Route principale
- Surface en eau



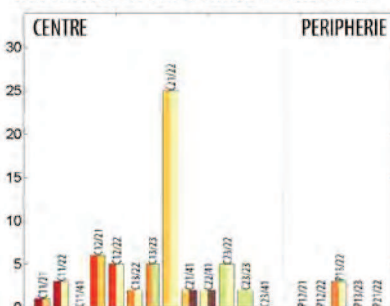
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE BESANCON

LEGENDE

MORPHOTYPES

- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

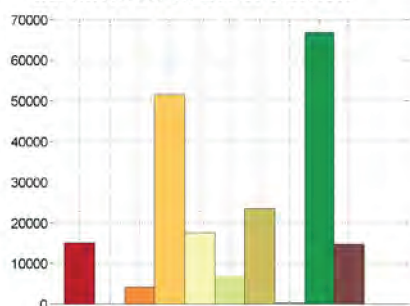
- Voie ferrée
- Route principale
- Surface en eau



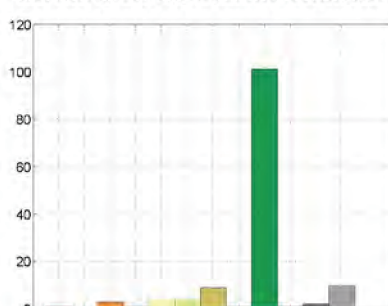
0 2 4 KM



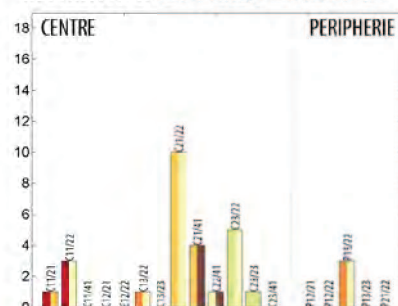
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE BÉTHUNE

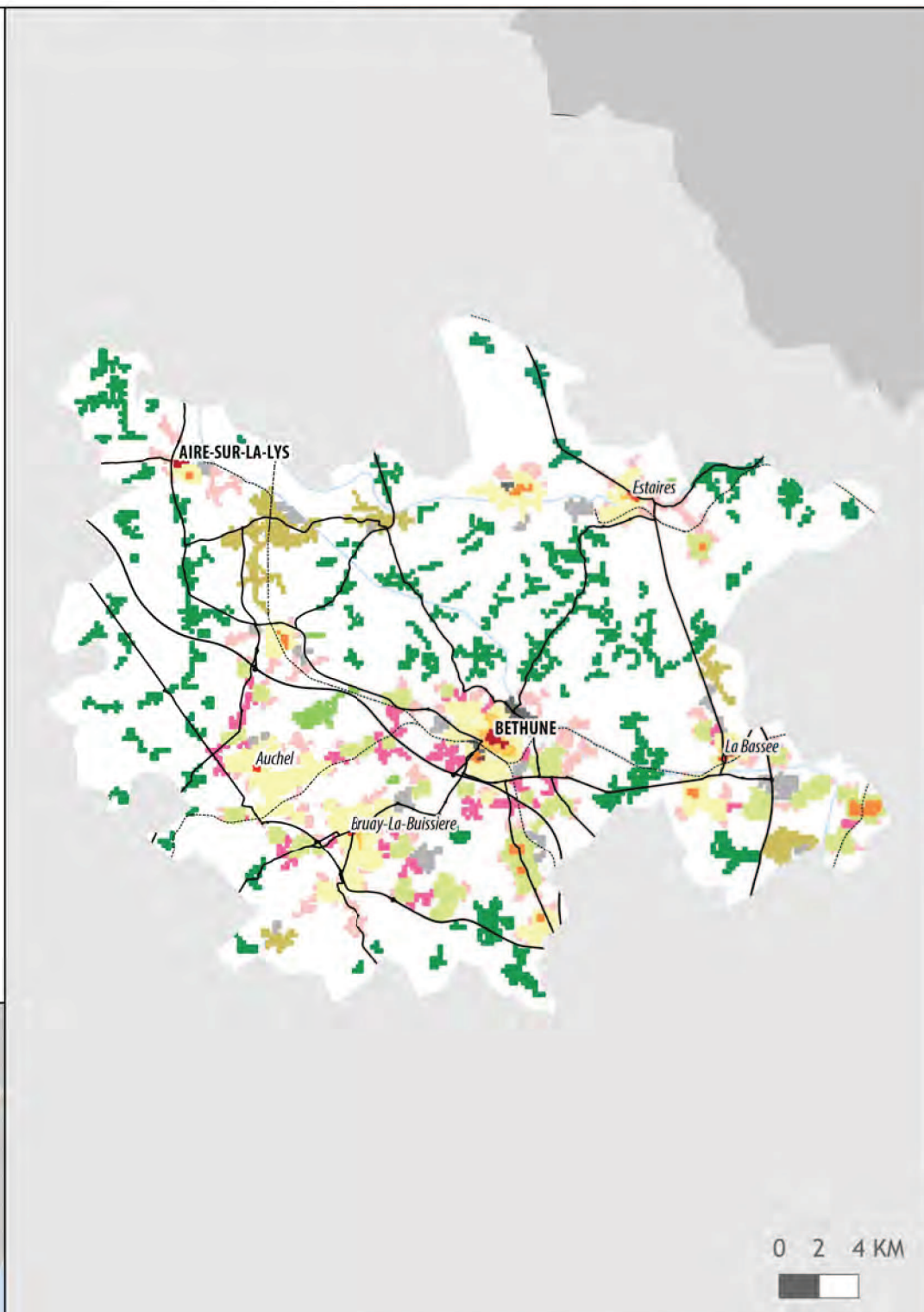
LEGENDE

MORPHOTYPES

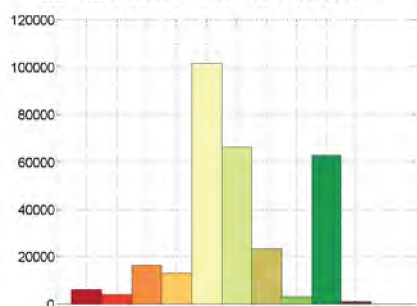
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

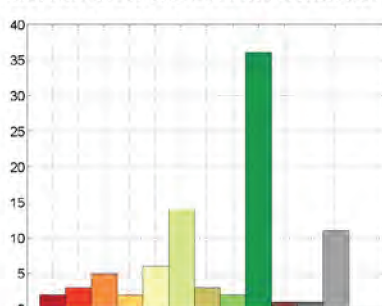
- Voie ferrée
- Route principale
- Surface en eau



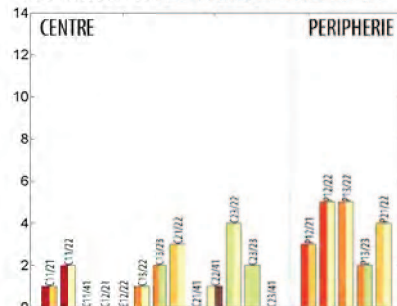
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE BORDEAUX

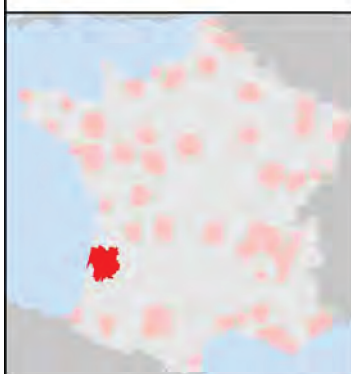
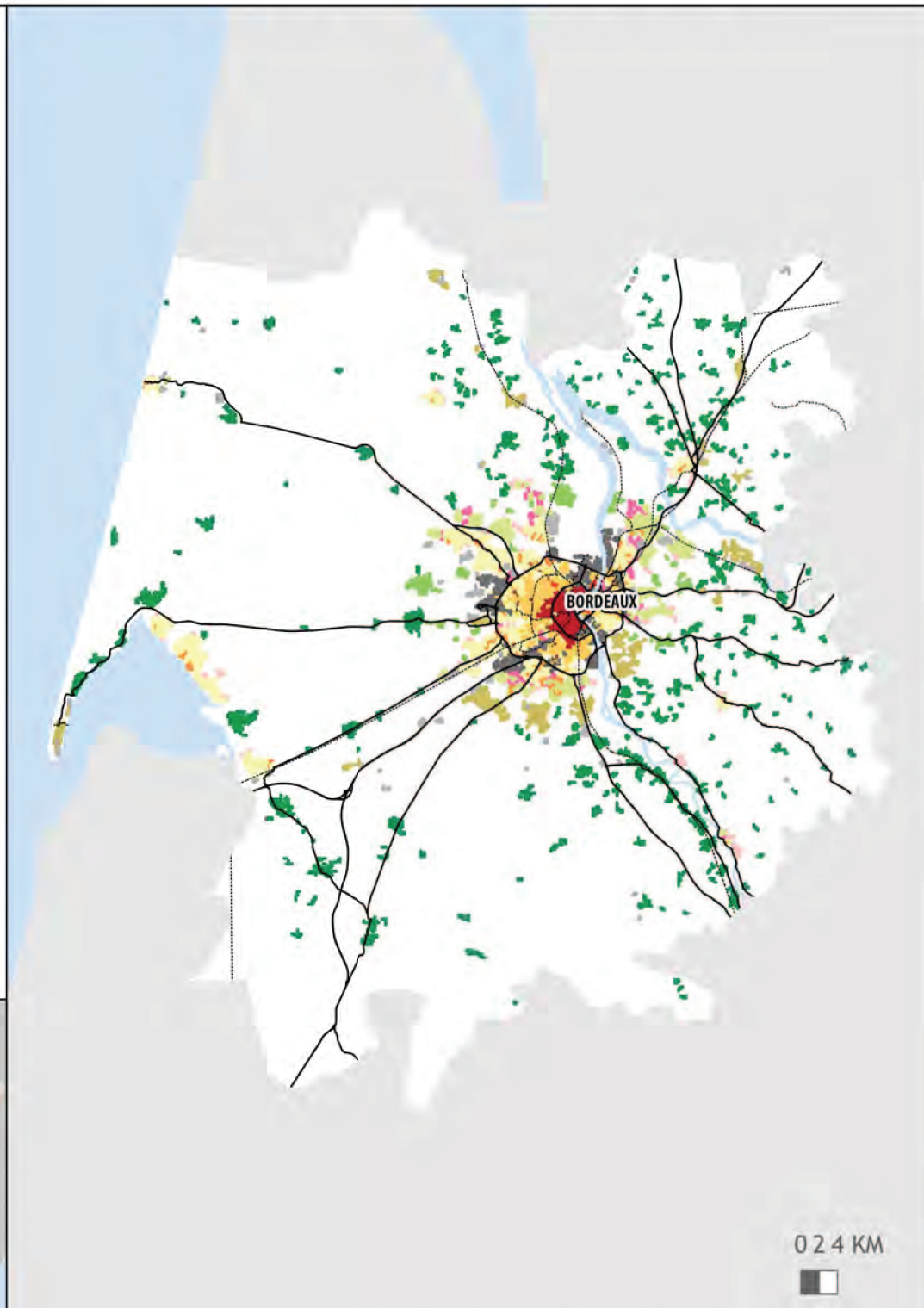
LEGENDE

MORPHOTYPES

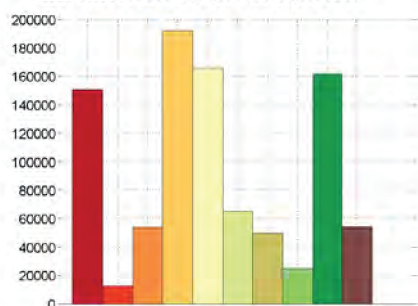
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

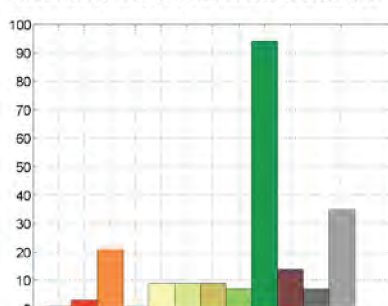
- Voie ferrée
- Route principale
- Surface en eau



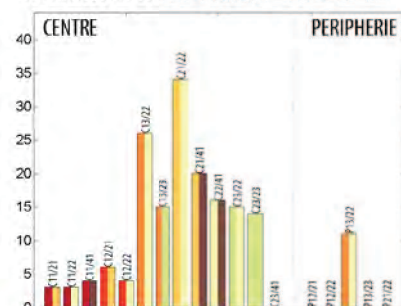
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE BREST

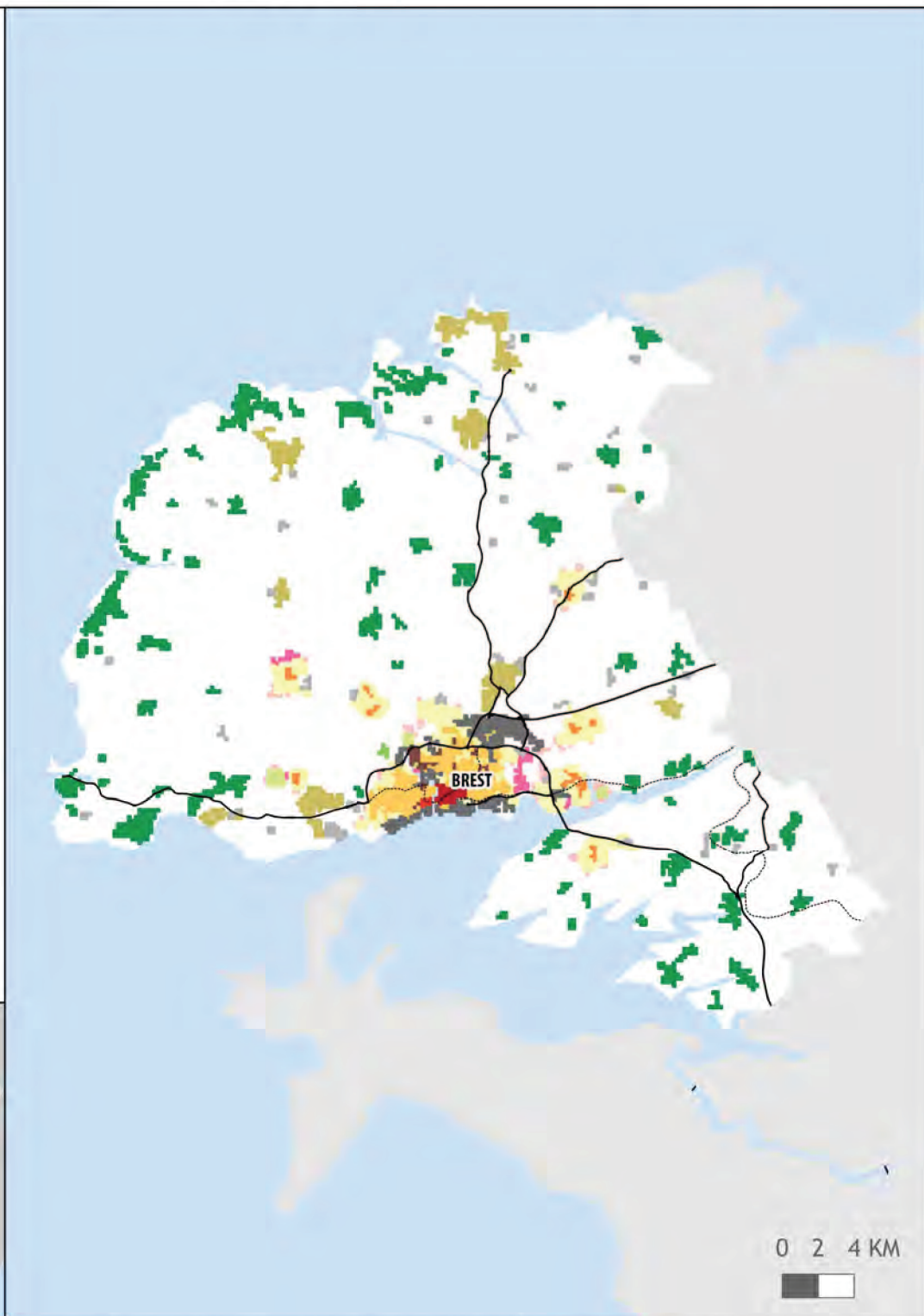
LEGENDE

MORPHOTYPES

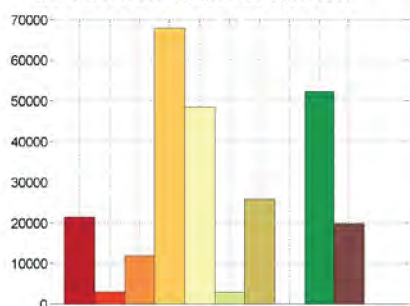
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

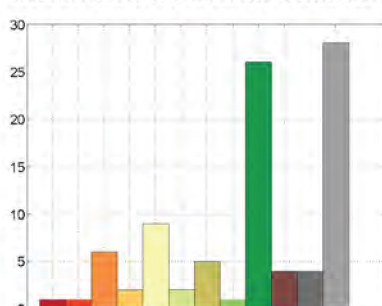
- Voie ferrée
- Route principale
- Surface en eau



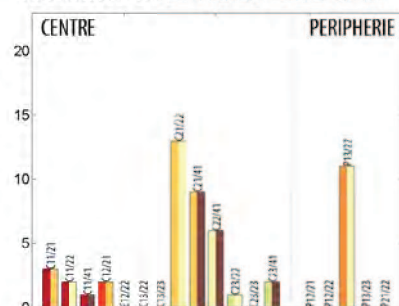
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE CAEN

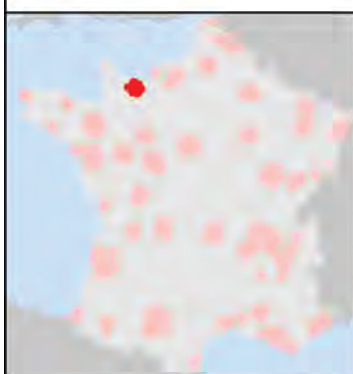
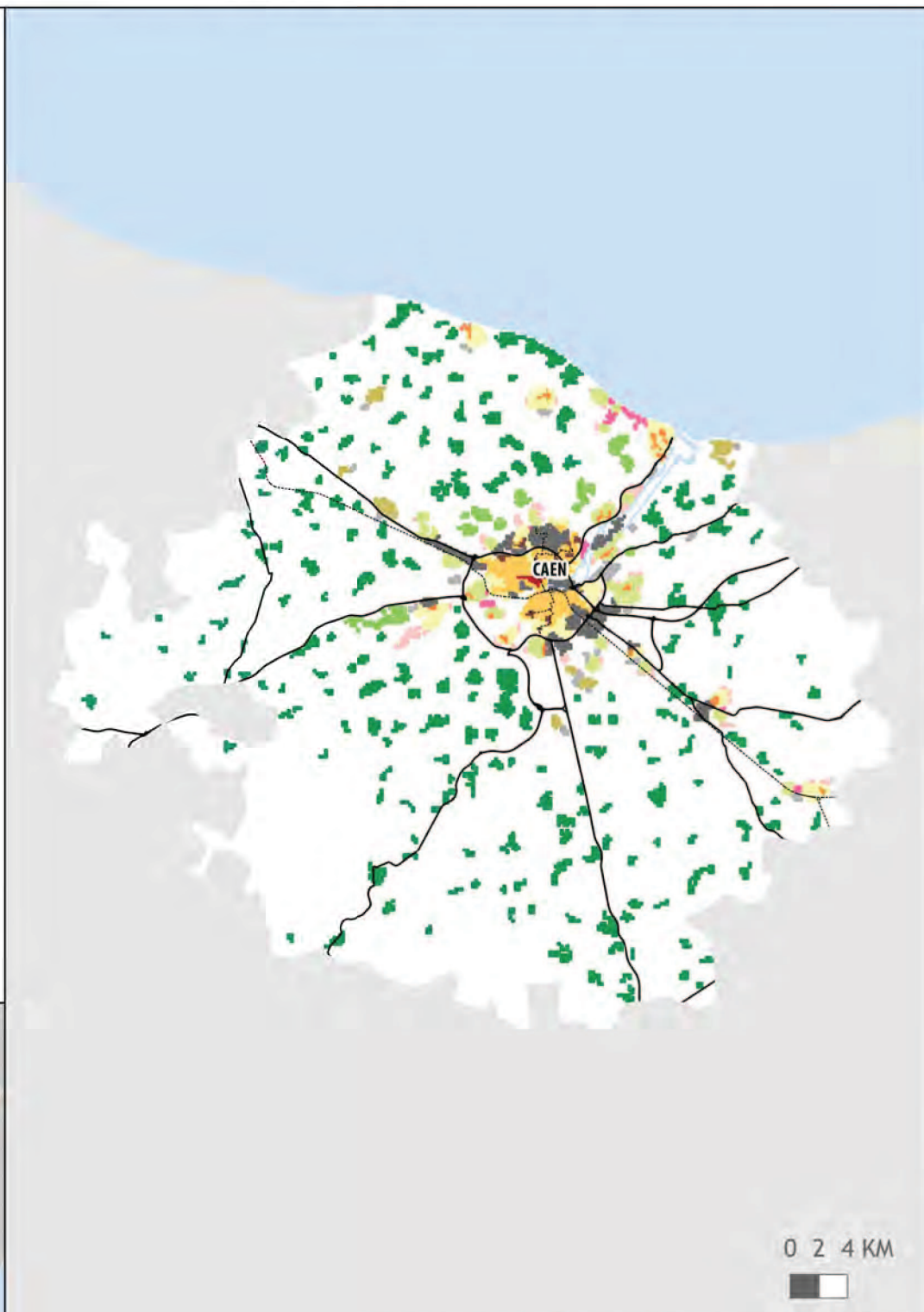
LEGENDE

MORPHOTYPES

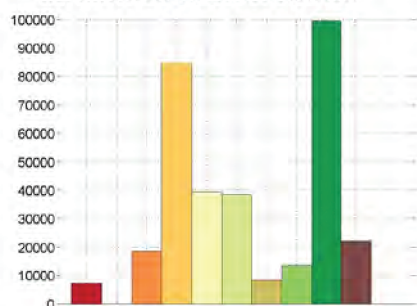
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

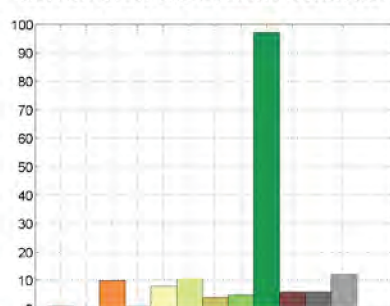
- Voie ferrée
- Route principale
- Surface en eau



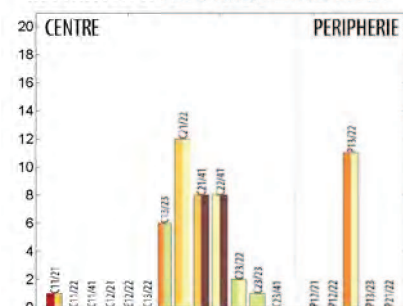
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE CHAMBERY

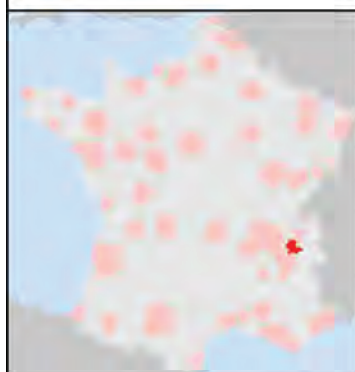
LEGENDE

MORPHOTYPES

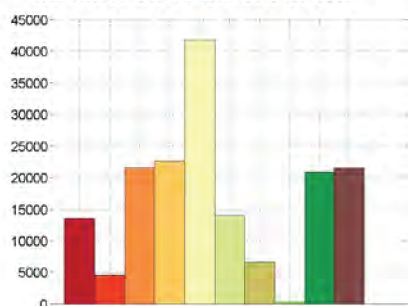
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

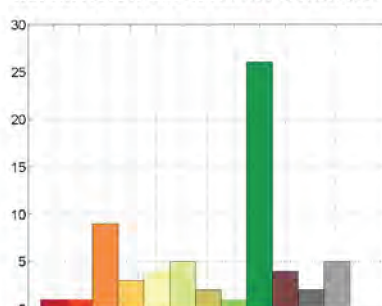
- Voie ferrée
- Route principale
- Surface en eau



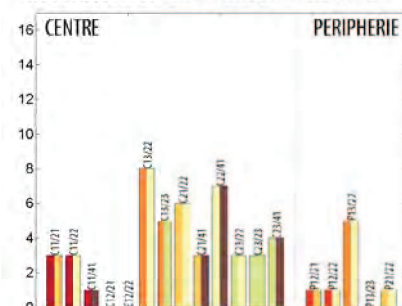
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE CLERMONT-FERRAND

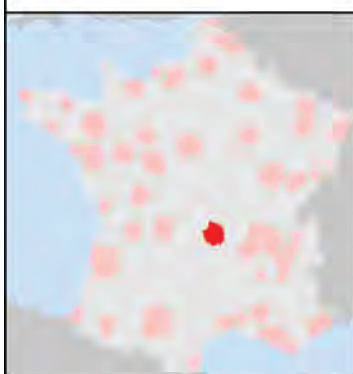
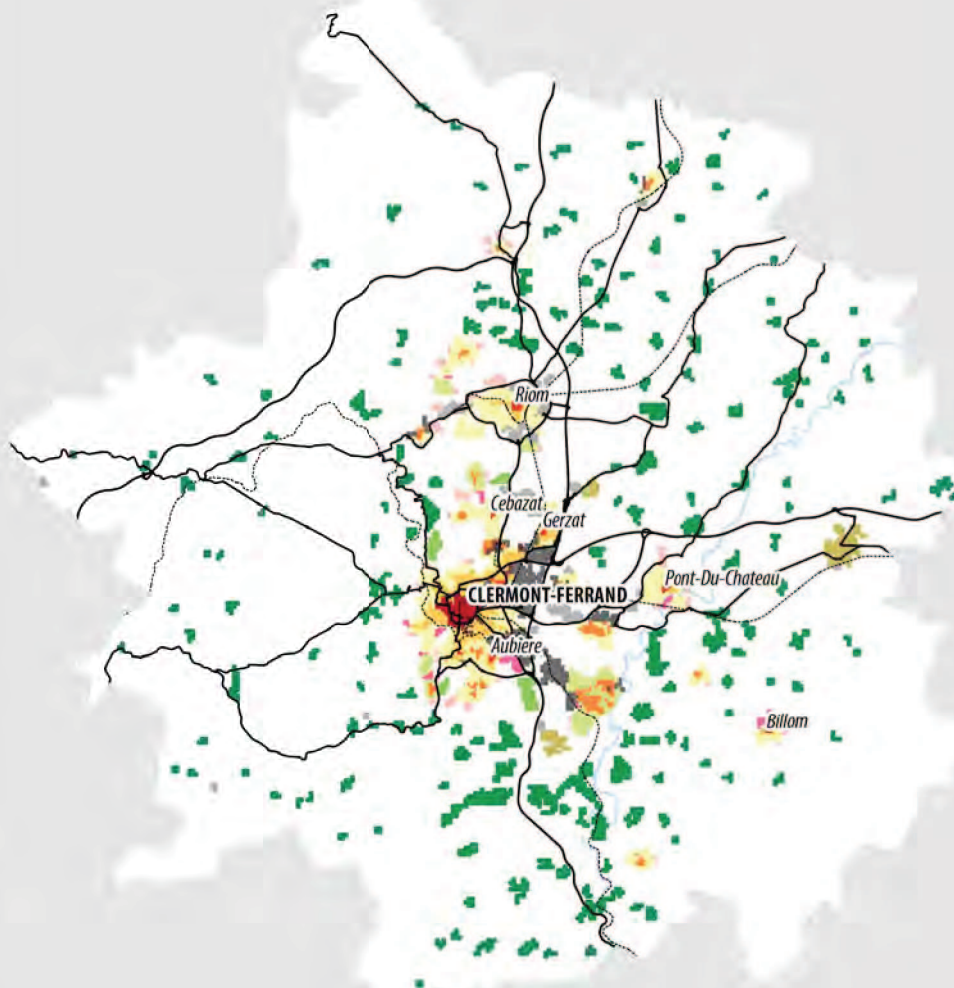
LEGENDE

MORPHOTYPES

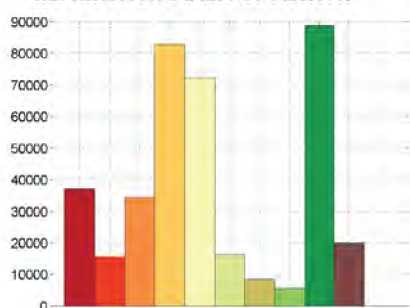
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

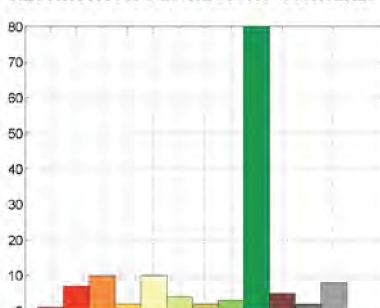
- Voie ferrée
- Route principale
- Surface en eau



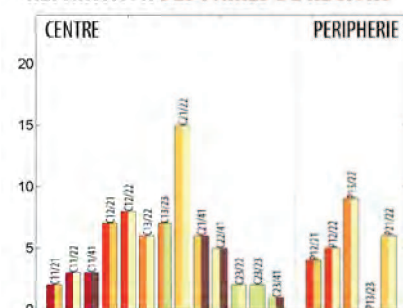
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE DIJON

LEGENDE

MORPHOTYPES

- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

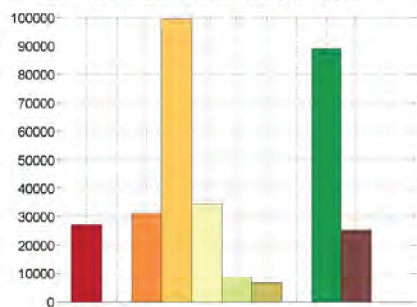
- Voie ferrée
- Route principale
- Surface en eau



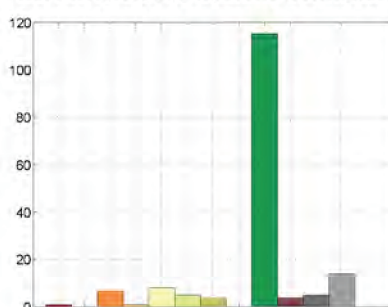
0 2 4 KM



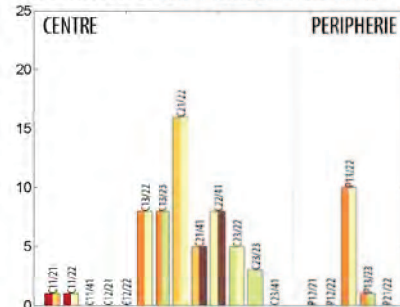
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE DOUAI - LENS

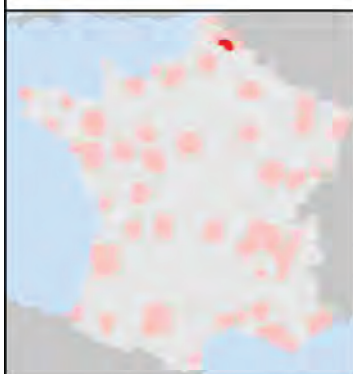
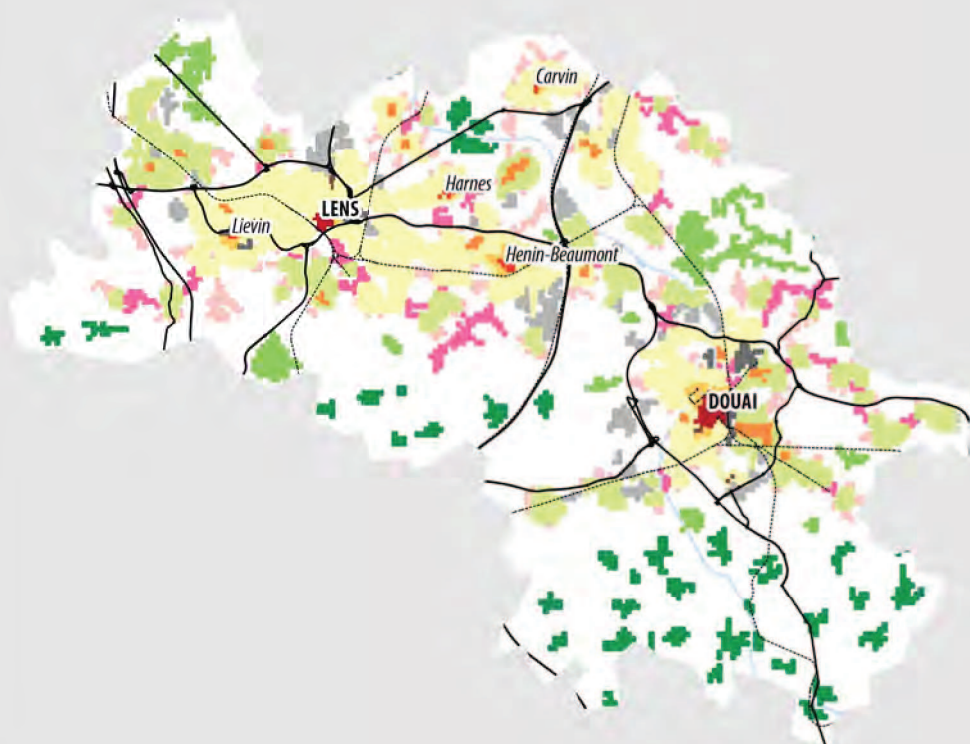
LEGENDE

MORPHOTYPES

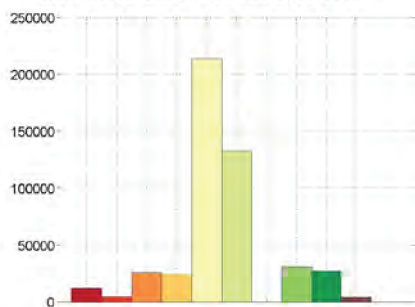
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

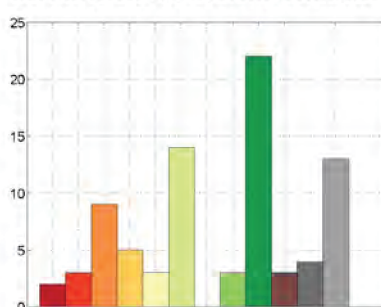
- Voie ferrée
- Route principale
- Surface en eau



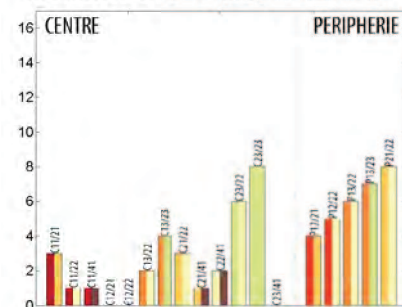
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE DUNKERQUE

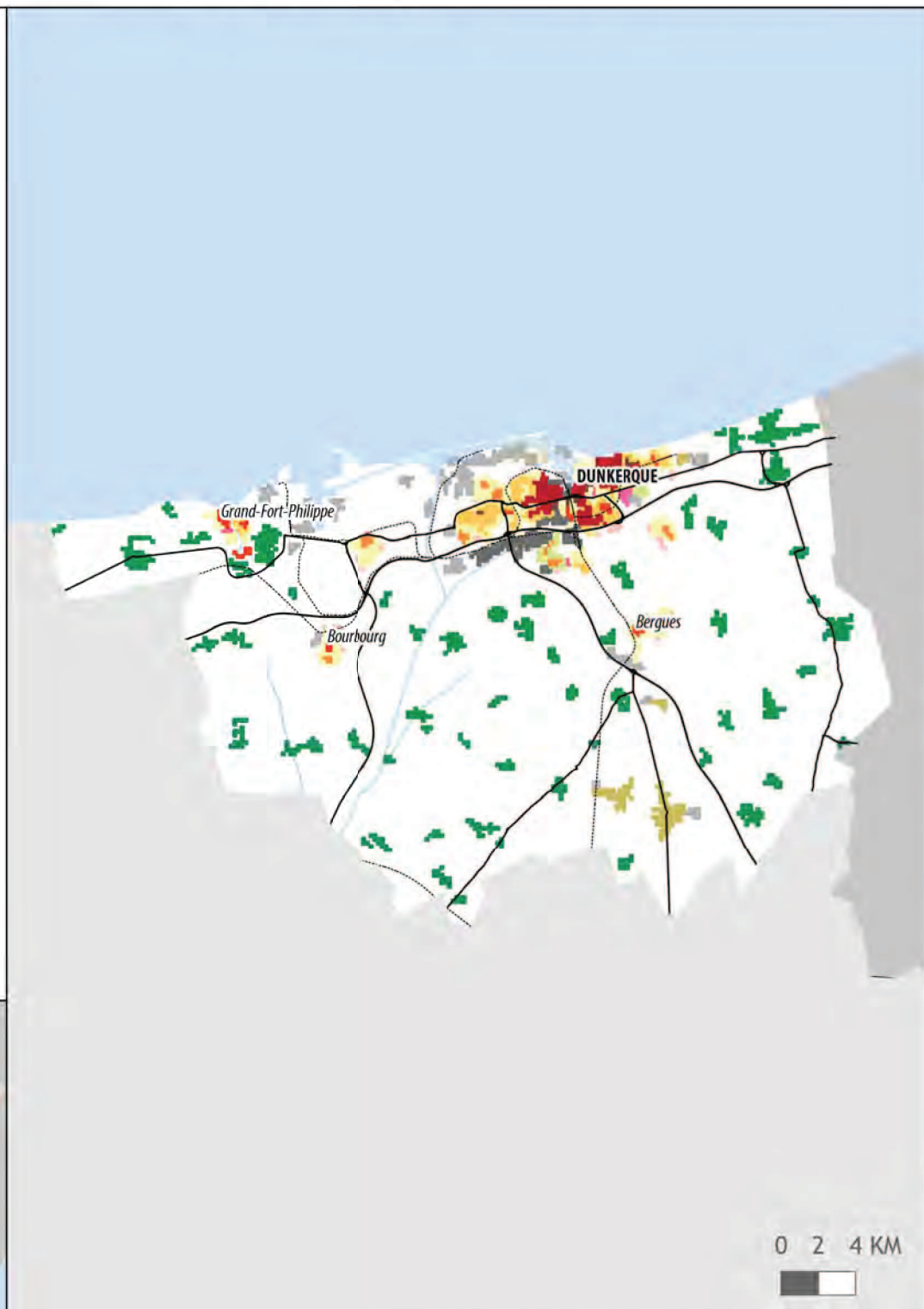
LEGENDE

MORPHOTYPES

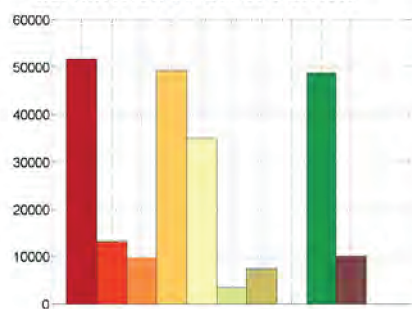
-  Centre principal
-  Centre secondaire
-  Centre tertiaire
-  Couronne dense
-  Couronne peu dense
-  Noyau connecté
-  Grands ensembles
-  Zone d'activité intégrée
-  Zone d'activité périphérique
-  Noyau relais
-  Noyau satellite
-  Noyau isolé
-  Zone connectrice
-  Frange urbaine

HABILLAGE

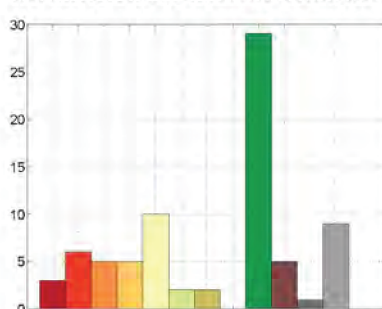
- Voie ferrée
— Route principale
Surface en eau



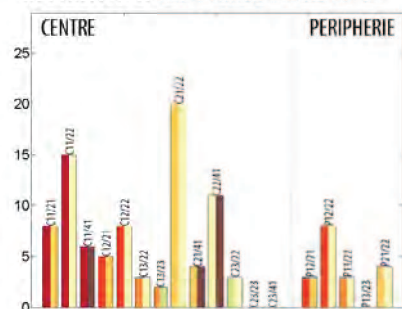
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE GRENOBLE

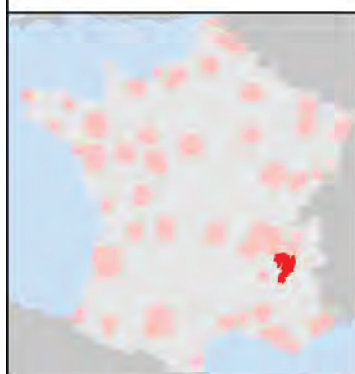
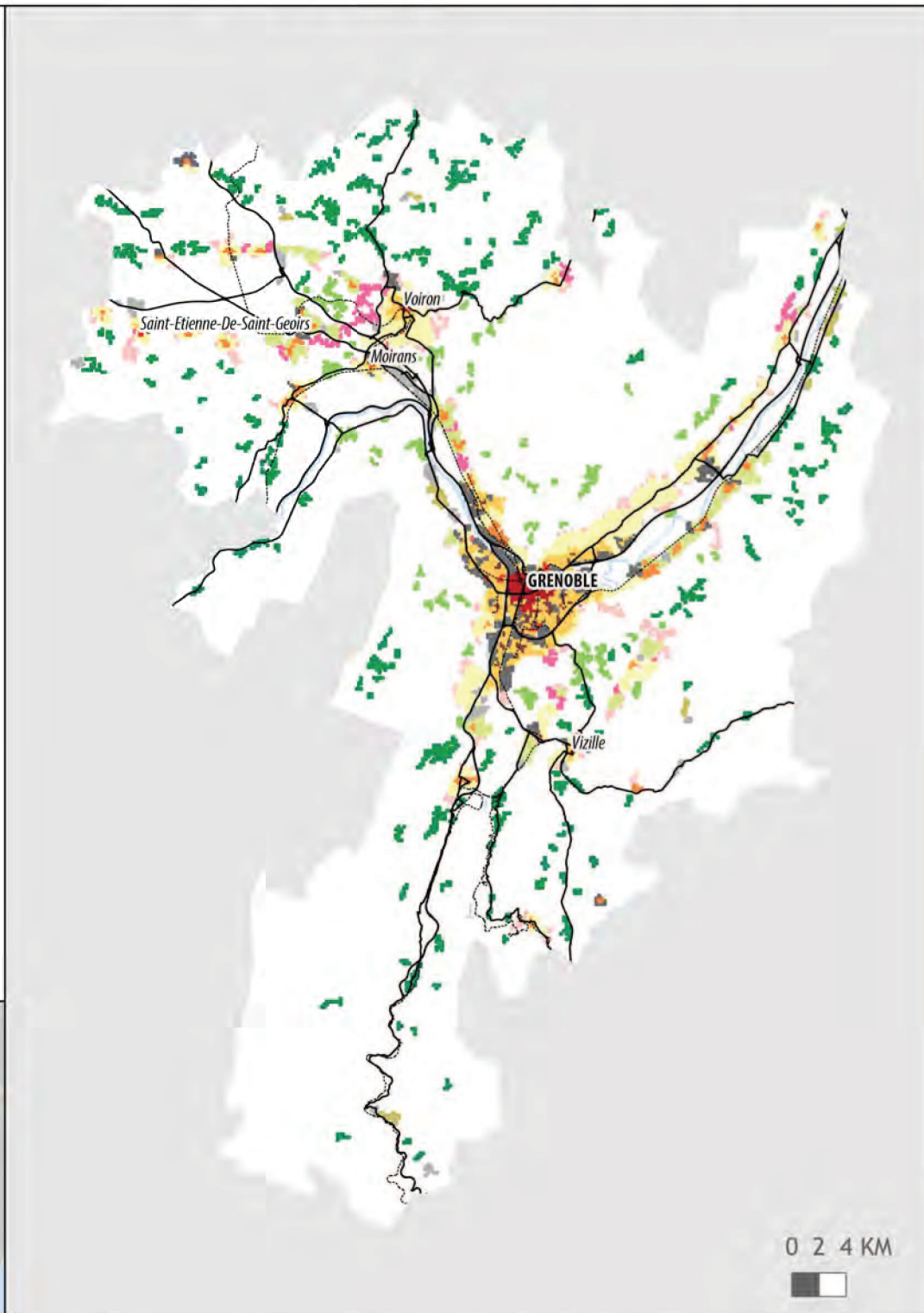
LEGENDE

MORPHOTYPES

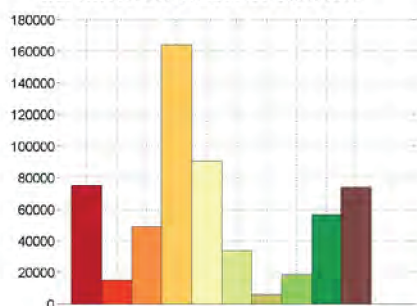
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

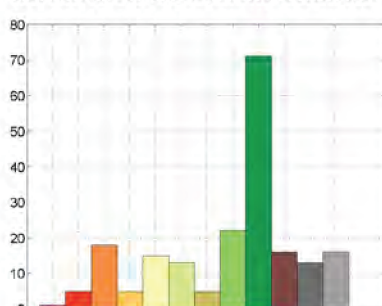
- Voie ferrée
- Route principale
- Surface en eau



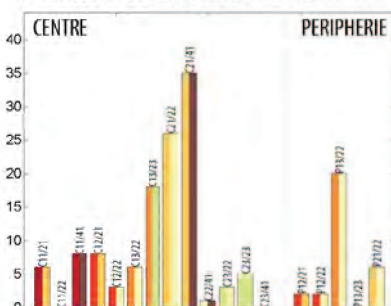
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE LA ROCHELLE

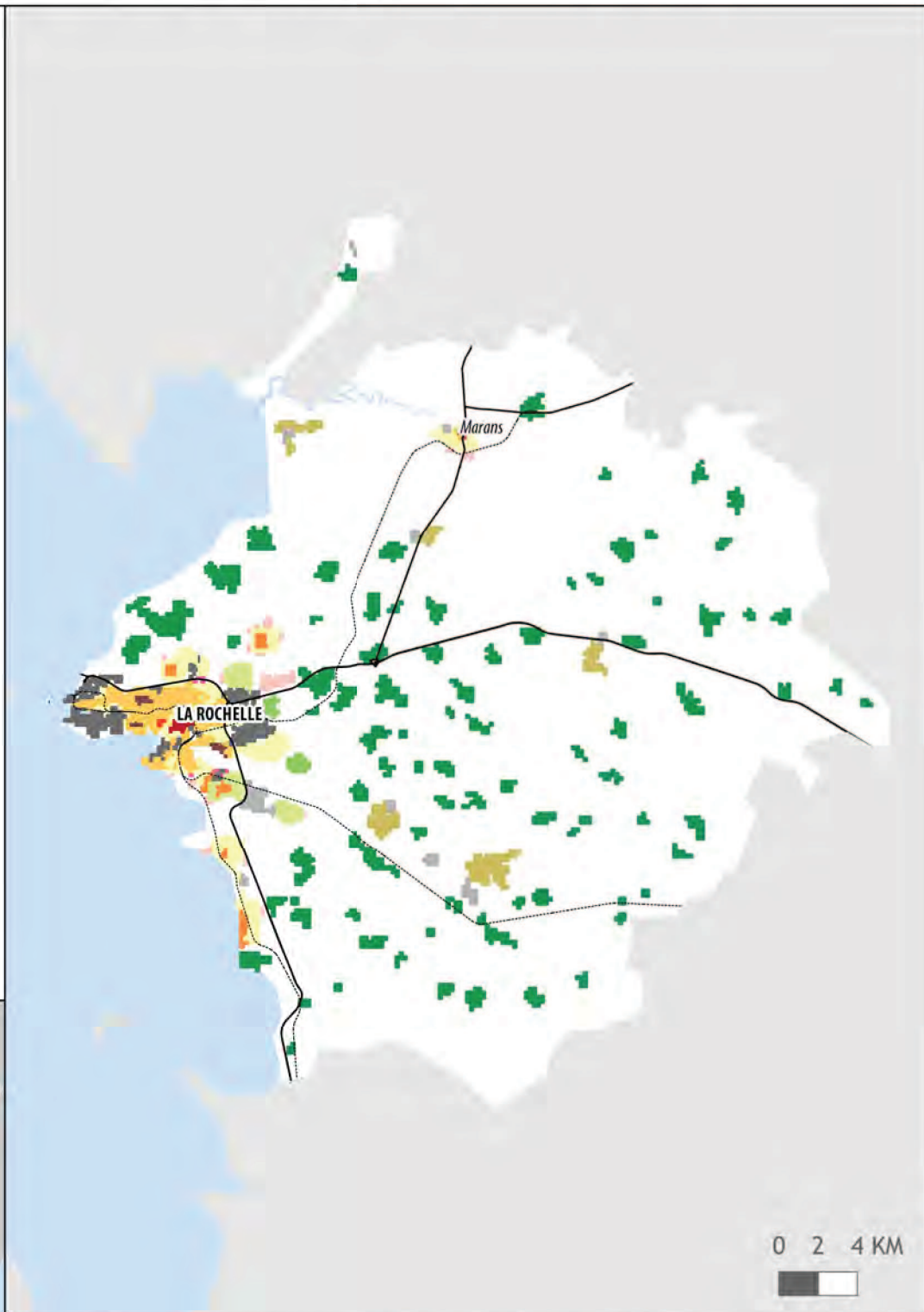
LEGENDE

MORPHOTYPES

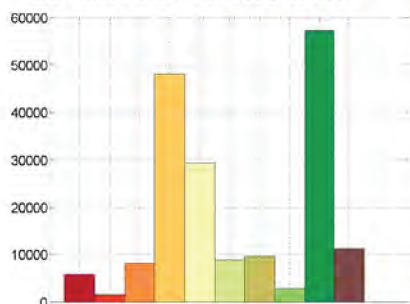
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

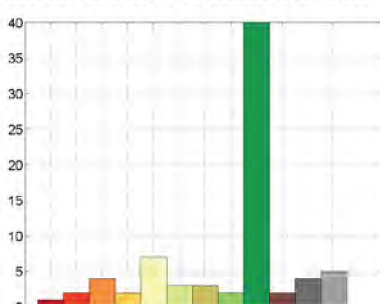
- Voie ferrée
- Route principale
- Surface en eau



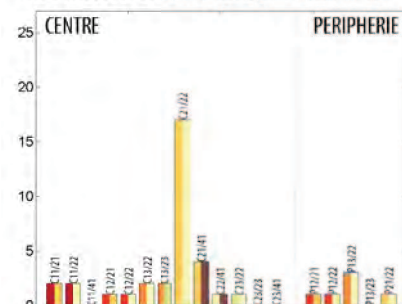
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DU HAVRE

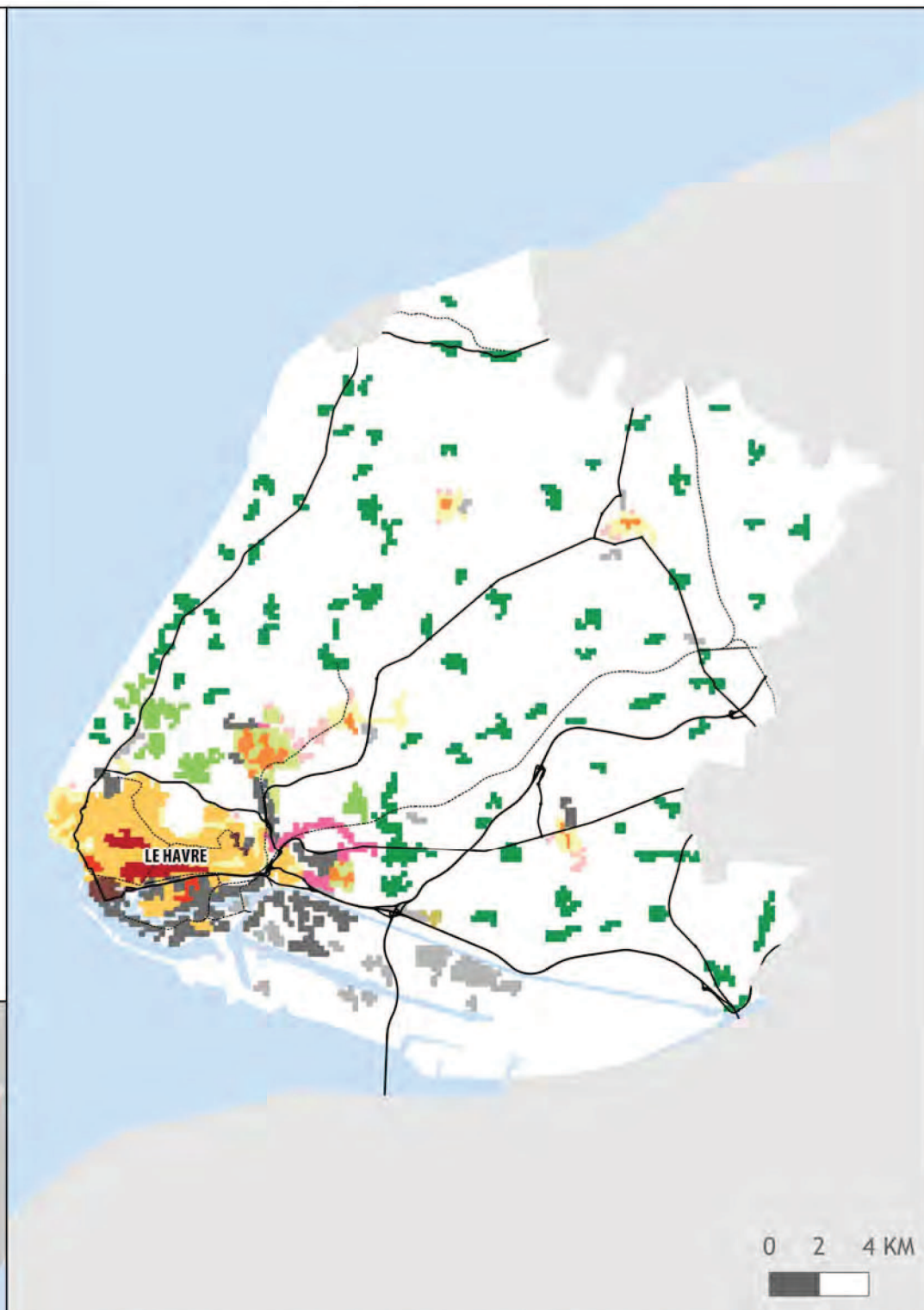
LEGENDE

MORPHOTYPES

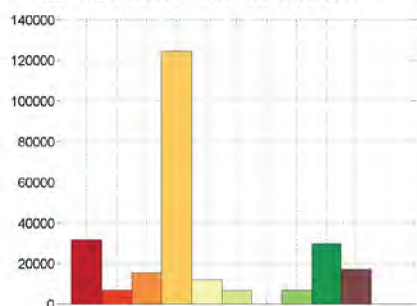
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

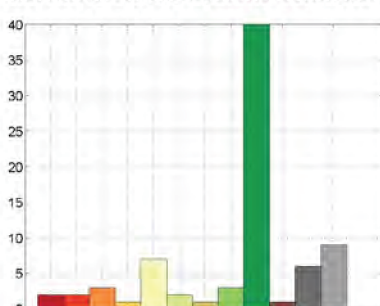
- Voie ferrée
- Route principale
- Surface en eau



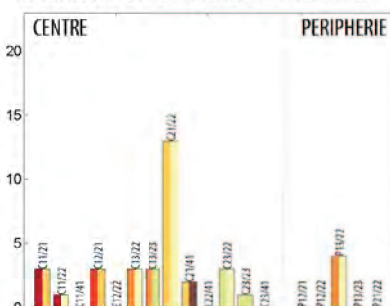
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DU MANS

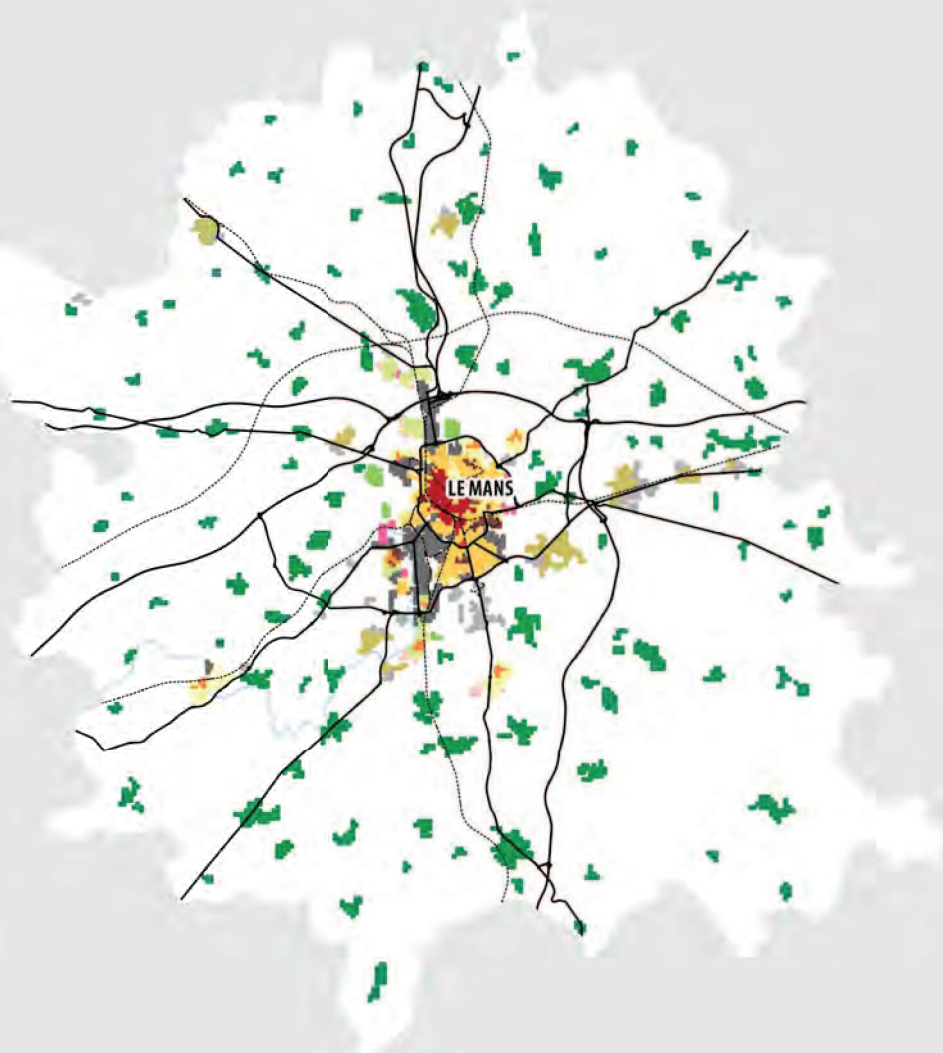
LEGENDE

MORPHOTYPES

- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

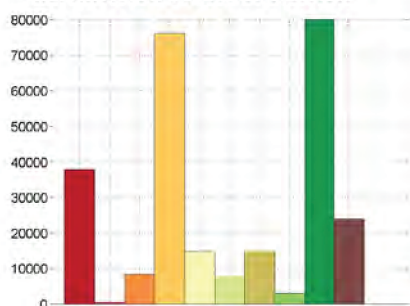
- Voie ferrée
- Route principale
- Surface en eau



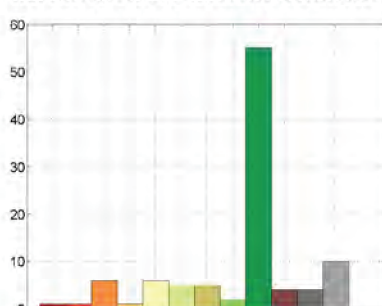
0 2 4 KM



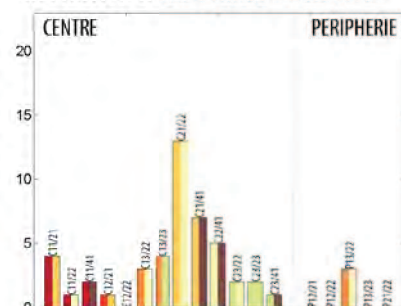
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE LILLE (PARTIE FRANÇAISE)

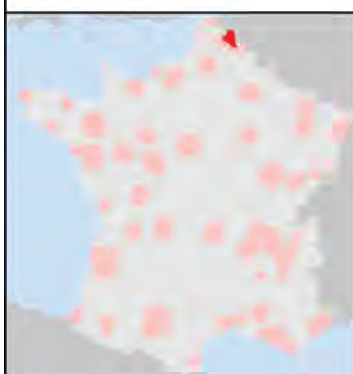
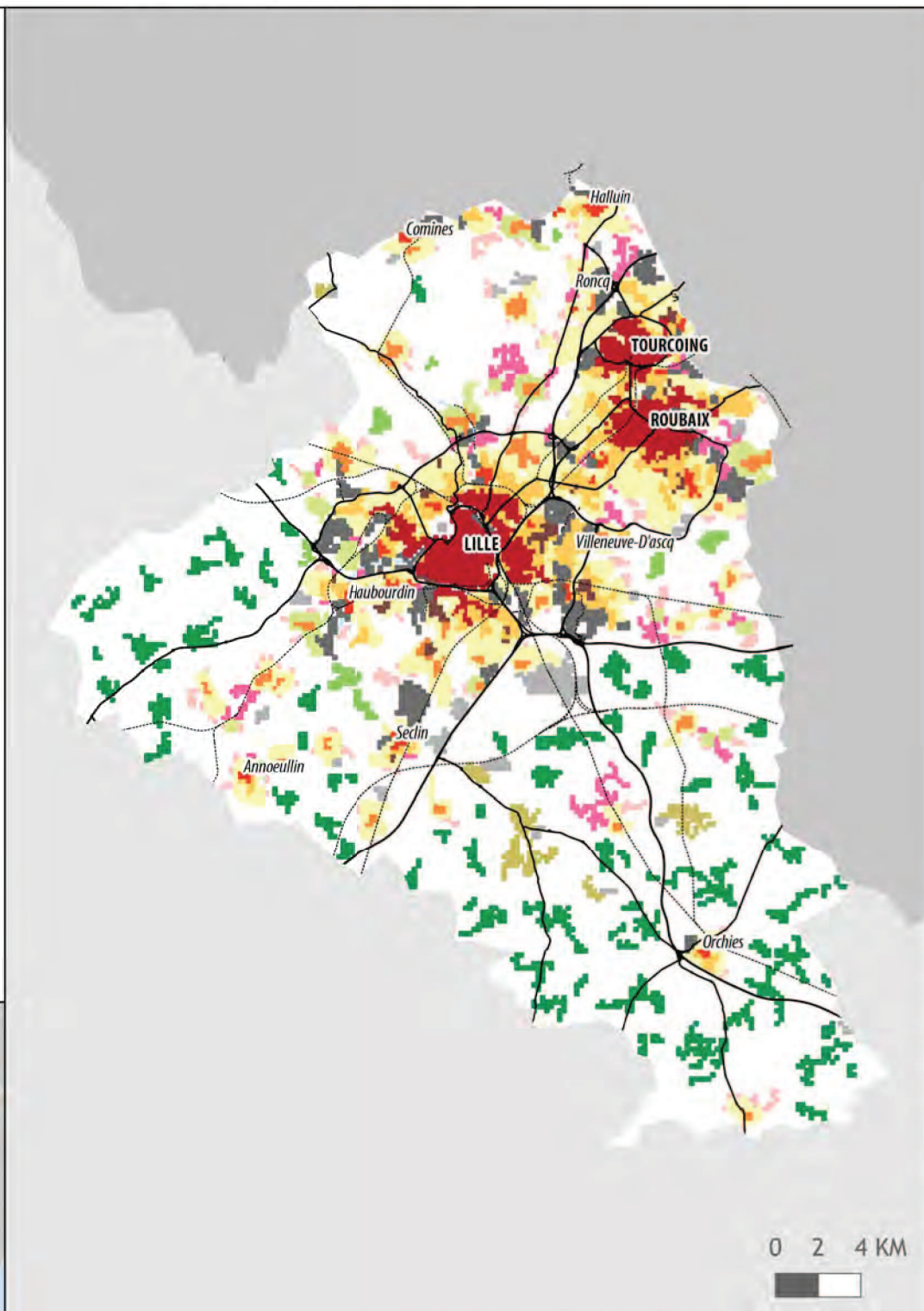
LEGENDE

MORPHOTYPES

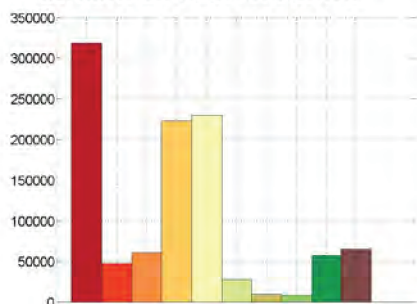
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

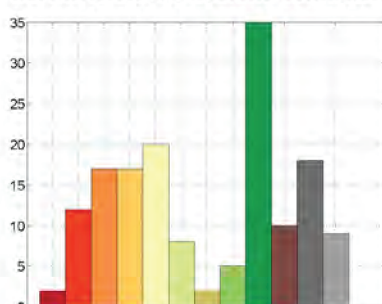
- Voie ferrée
- Route principale
- Surface en eau



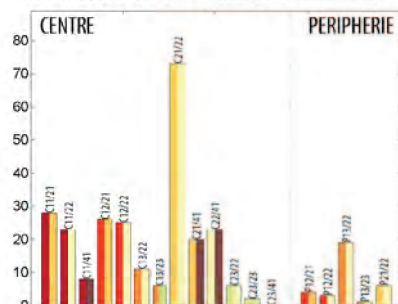
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE LIMOGES

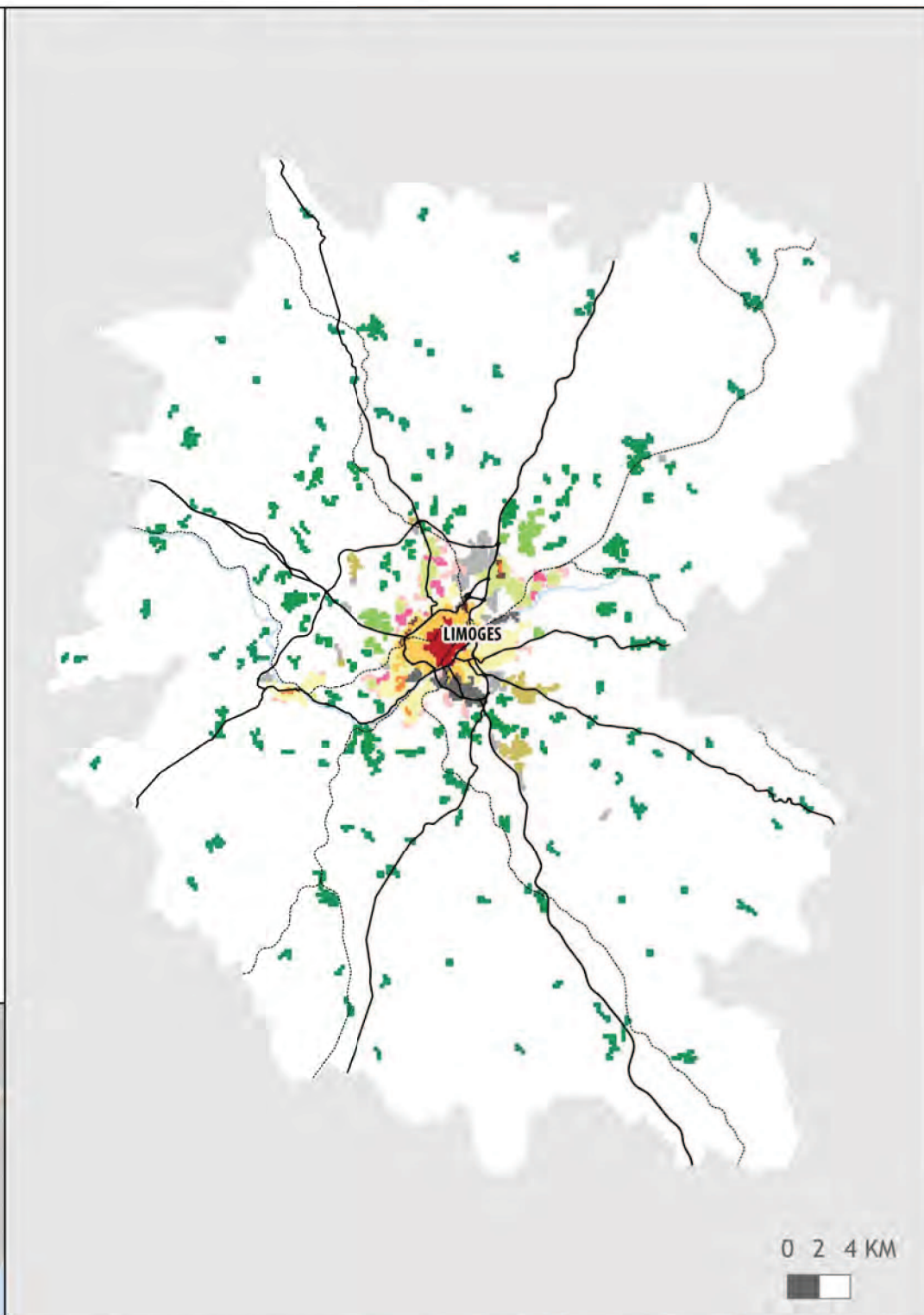
LEGENDE

MORPHOTYPES

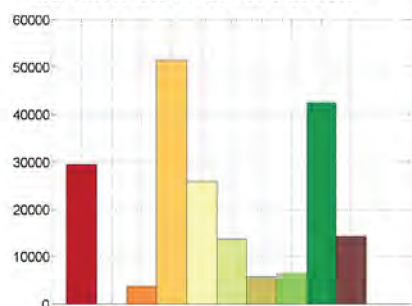
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

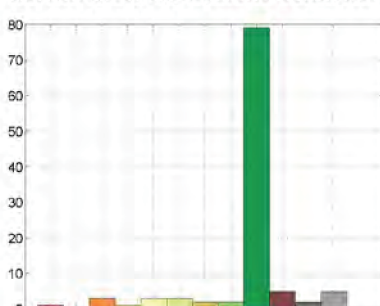
- Voie ferrée
- Route principale
- Surface en eau



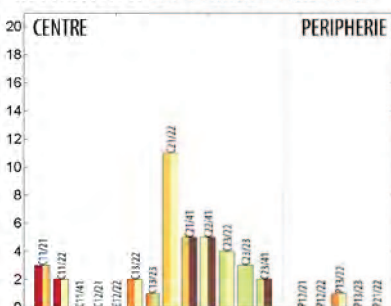
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE LORIENT

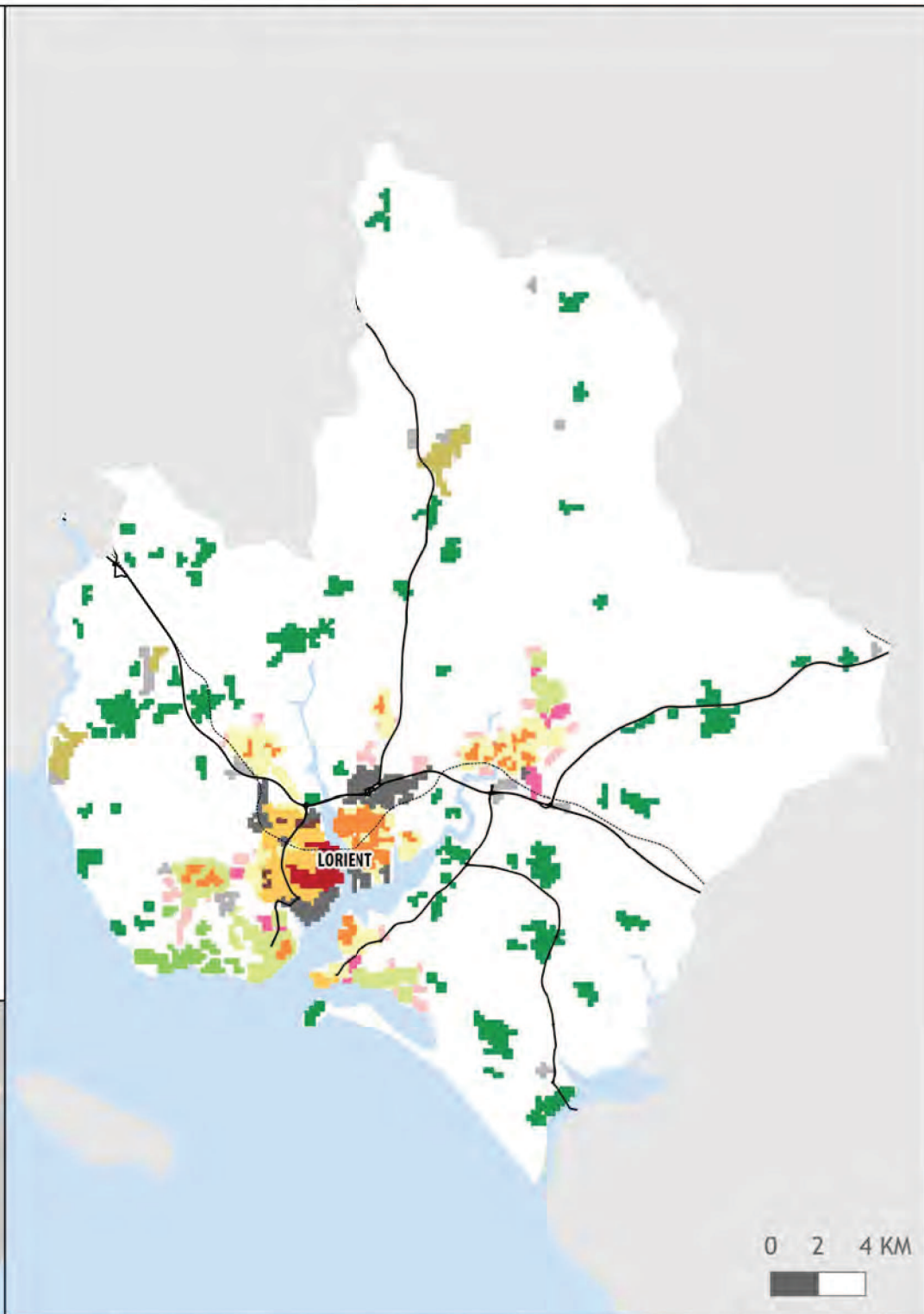
LEGENDE

MORPHOTYPES

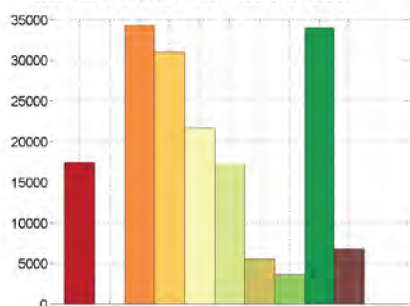
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

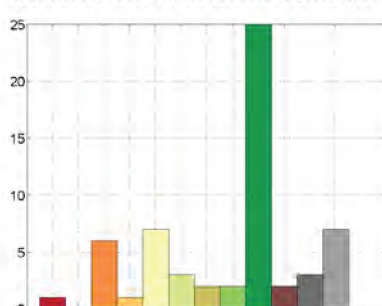
- Voie ferrée
- Route principale
- Surface en eau



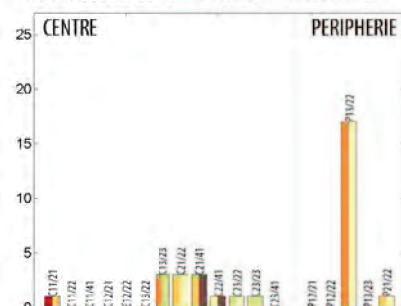
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE LYON

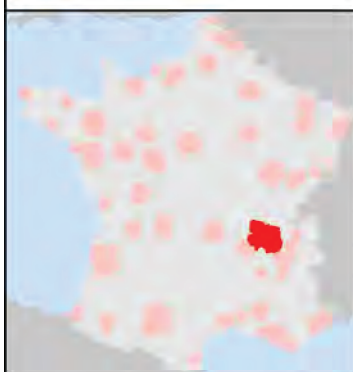
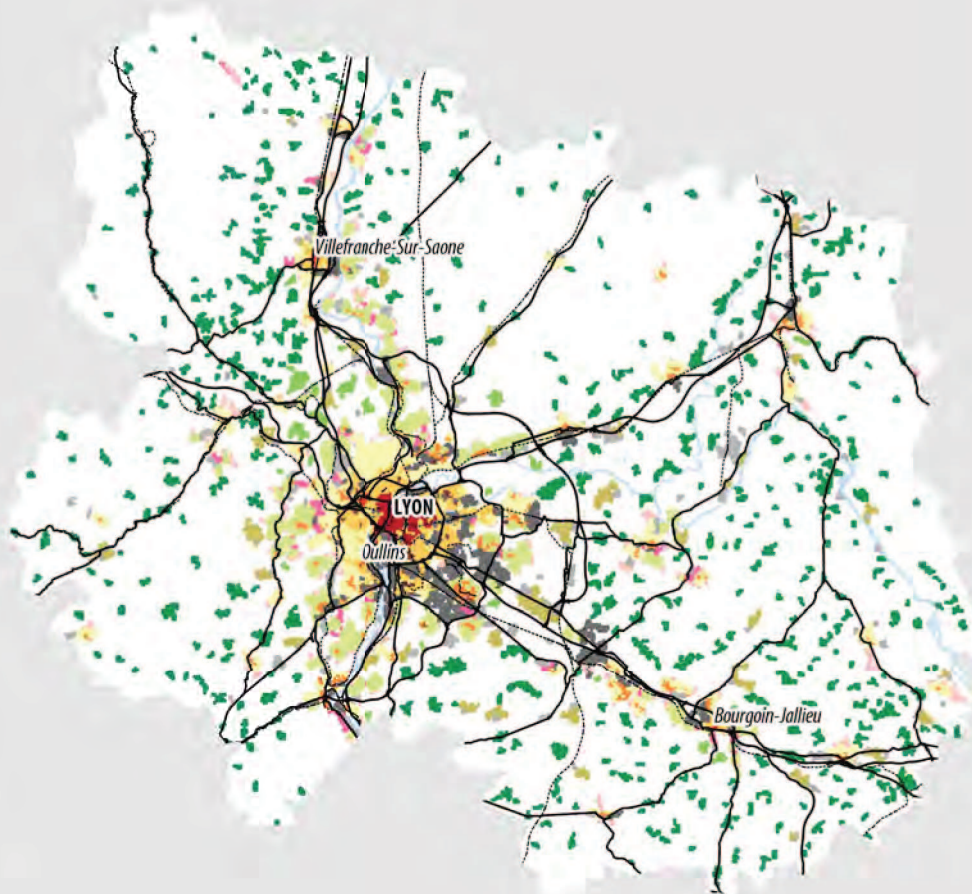
LEGENDE

MORPHOTYPES

- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

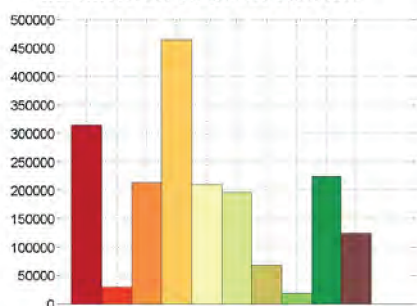
HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

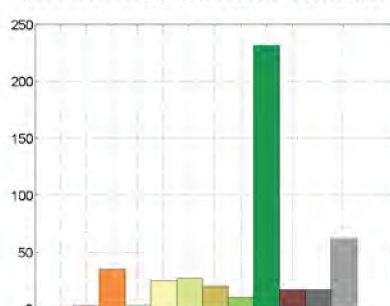


0.24 KM

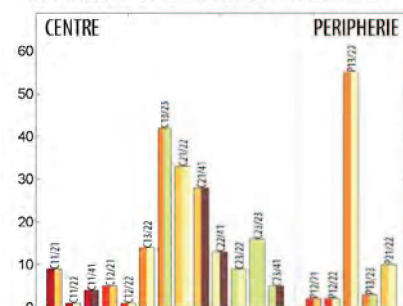
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE MARSEILLE - AIX-EN-PROVENCE

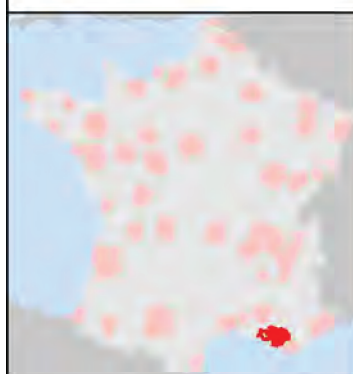
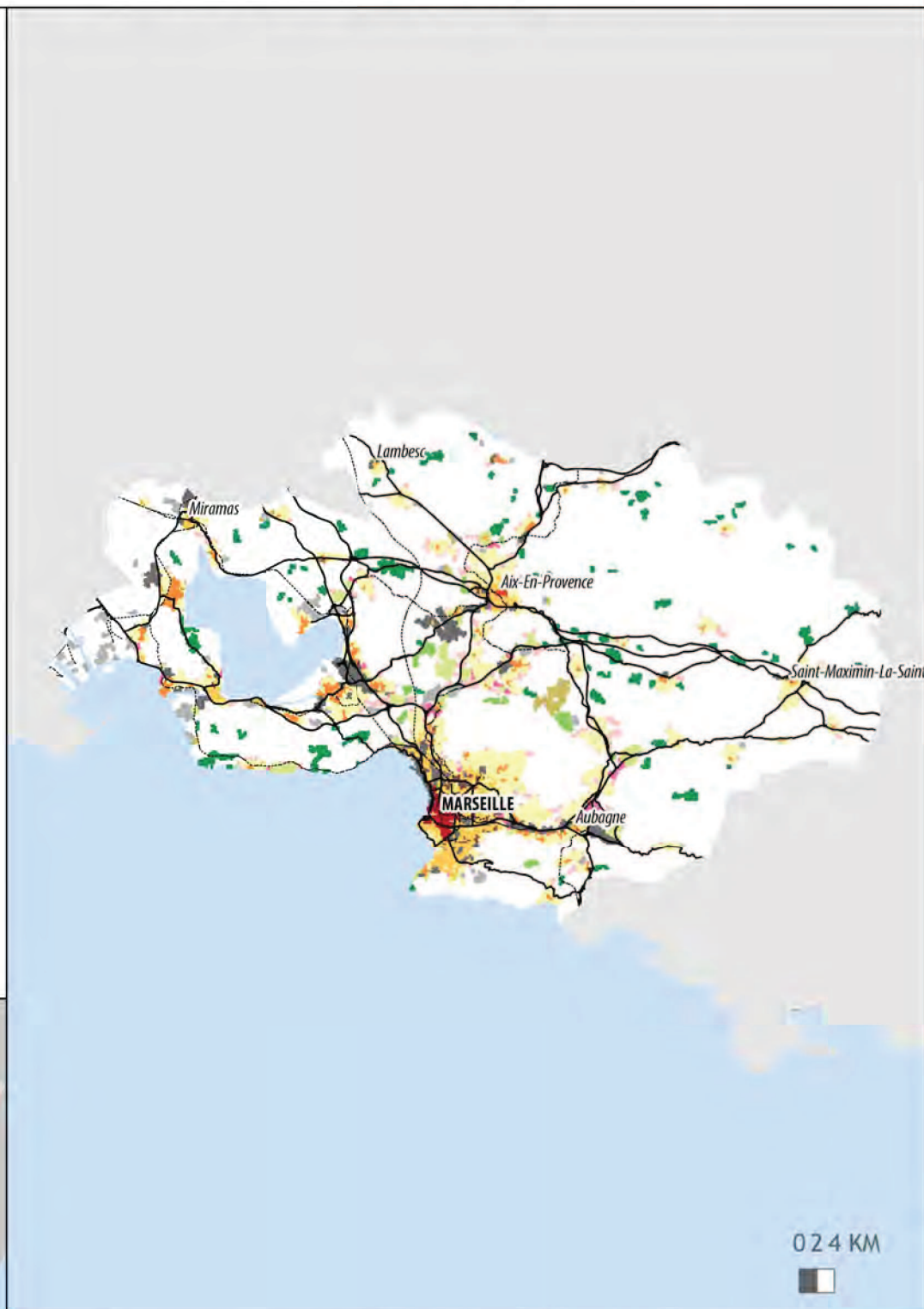
LEGENDE

MORPHOTYPES

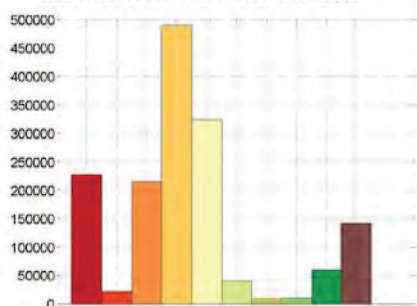
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

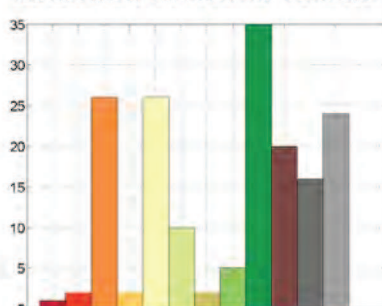
- Voie ferrée
- Route principale
- Surface en eau



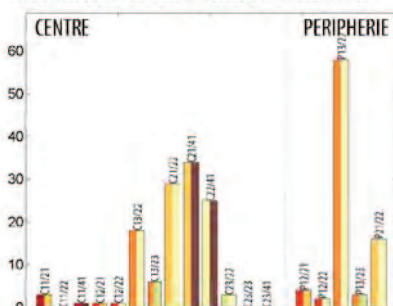
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE METZ

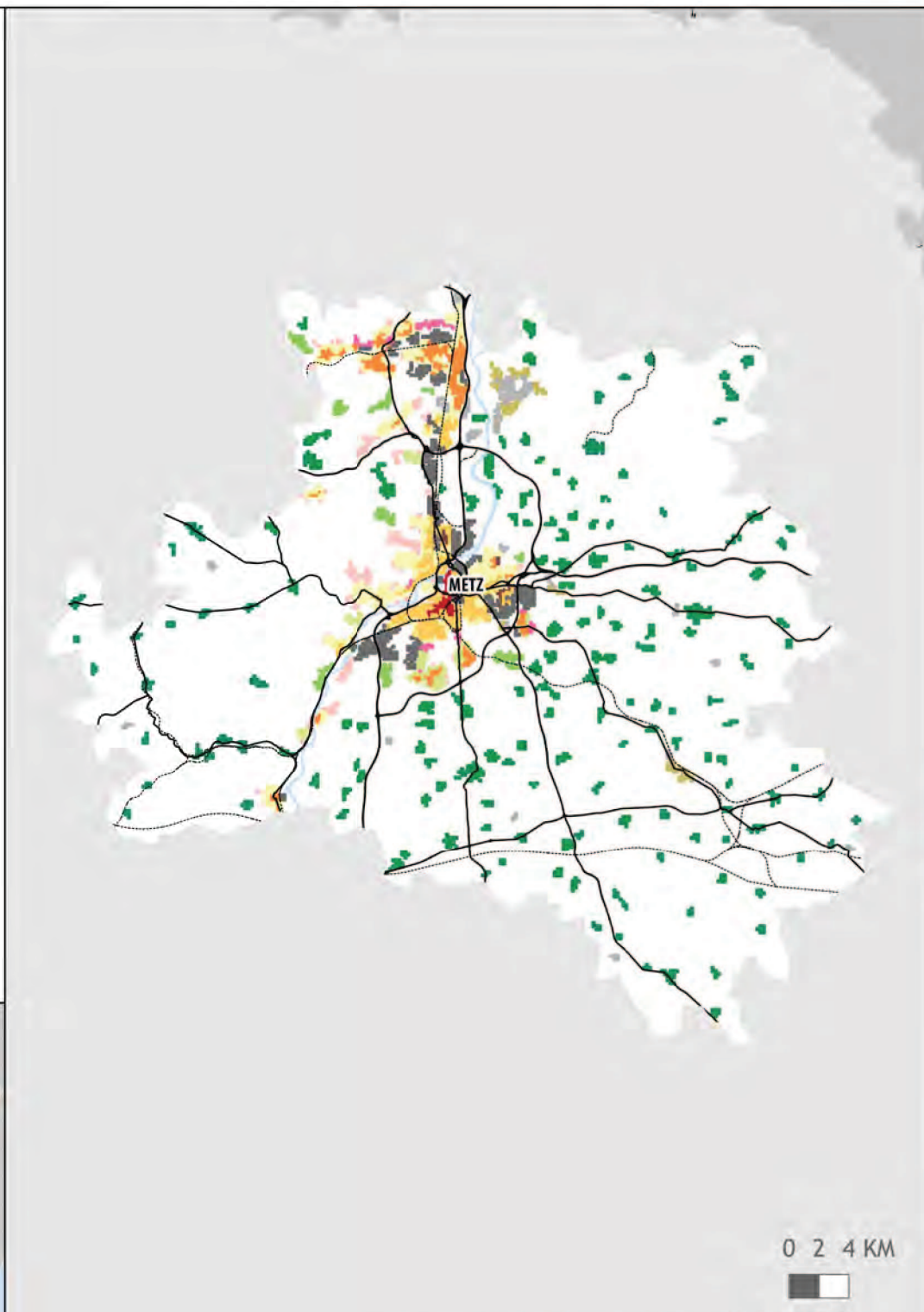
LEGENDE

MORPHOTYPES

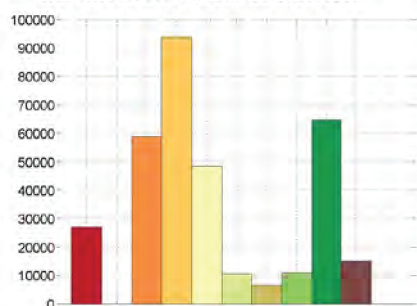
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

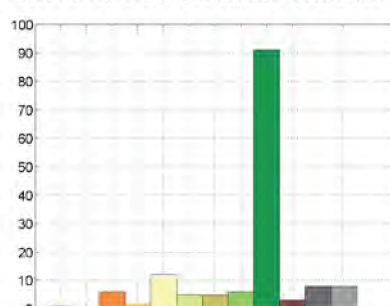
- Voie ferrée
- Route principale
- Surface en eau



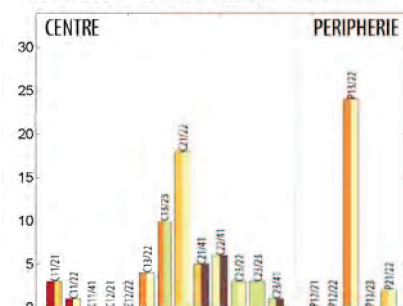
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE MONTBÉLIARD

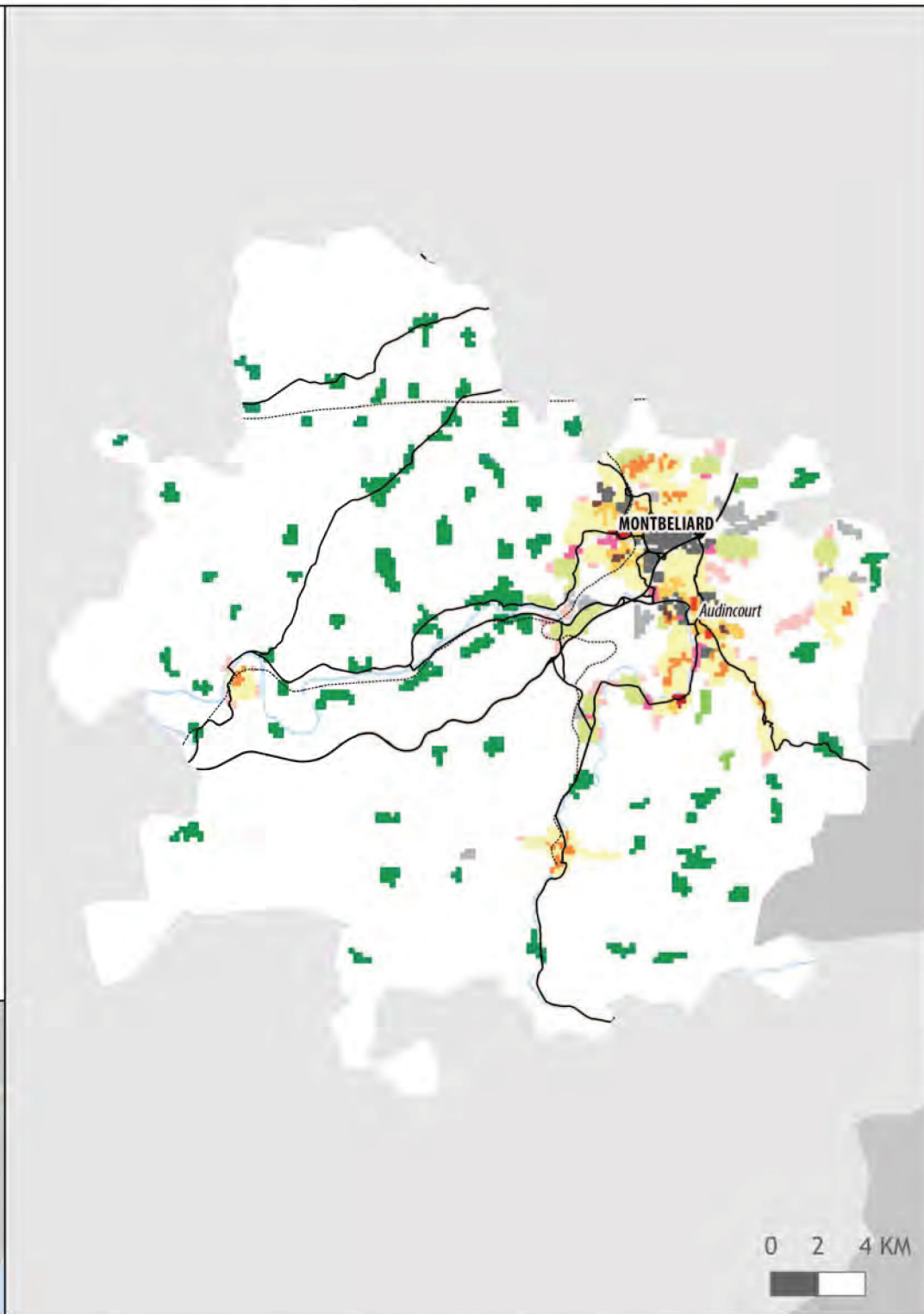
LEGENDE

MORPHOTYPES

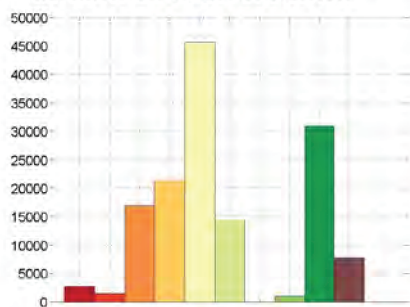
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

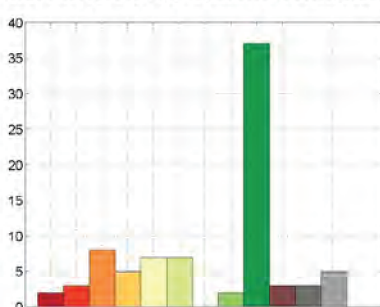
- Voie ferrée
- Route principale
- Surface en eau



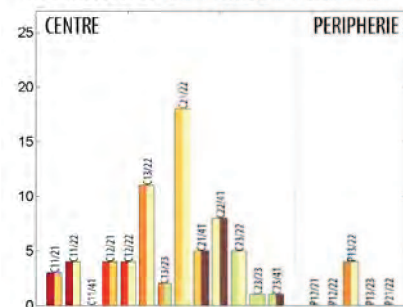
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE MONTPELLIER

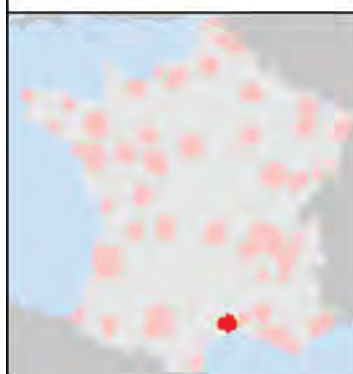
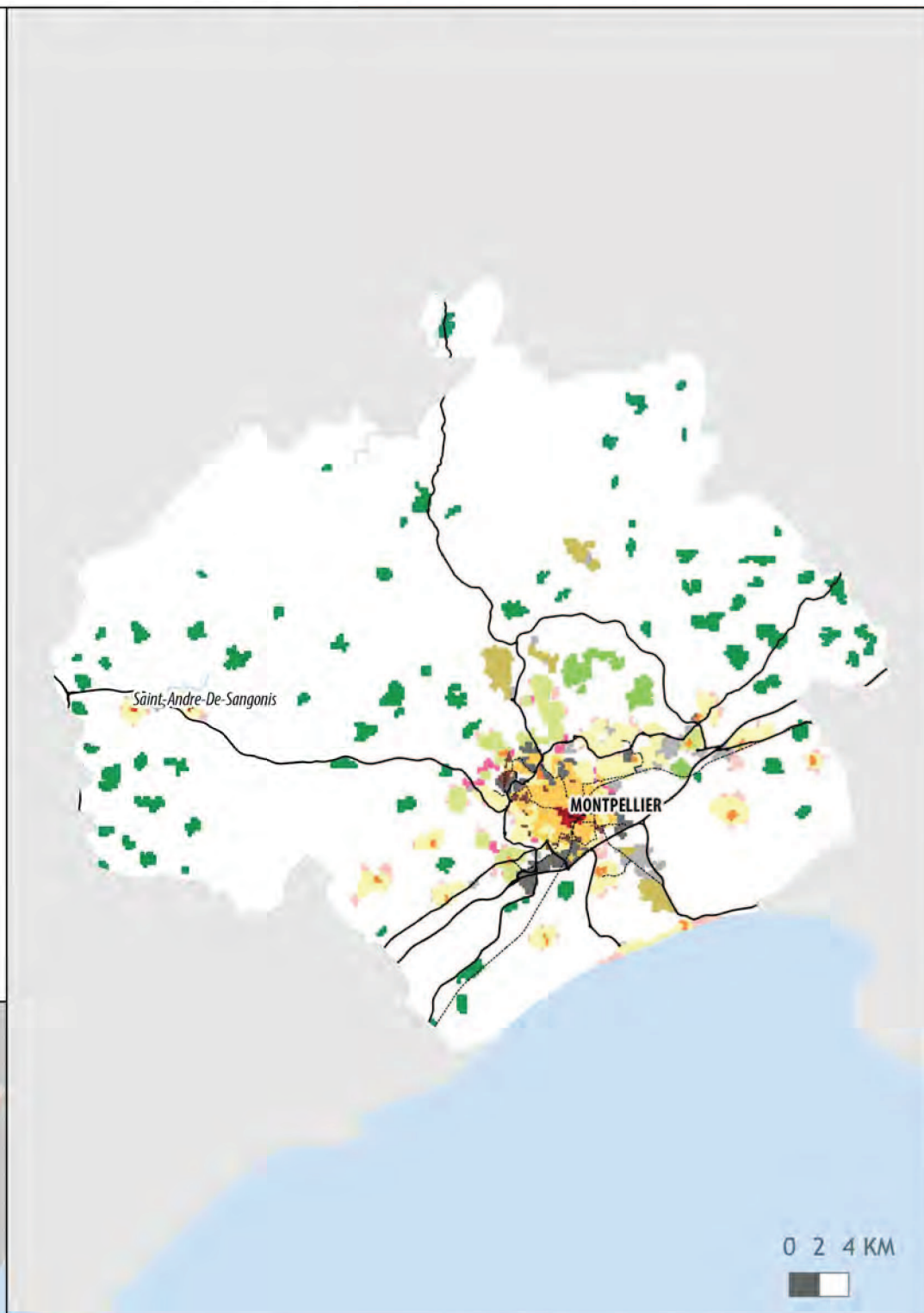
LEGENDE

MORPHOTYPES

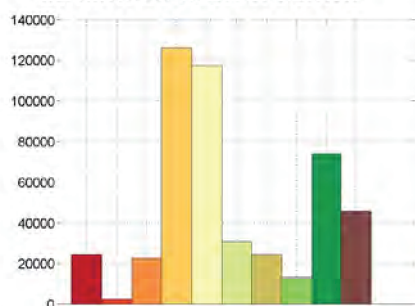
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

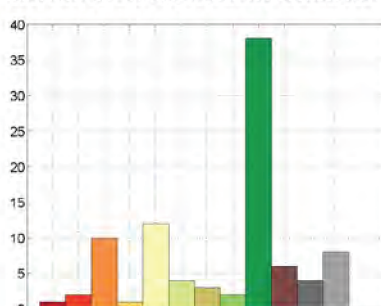
- Voie ferrée
- Route principale
- Surface en eau



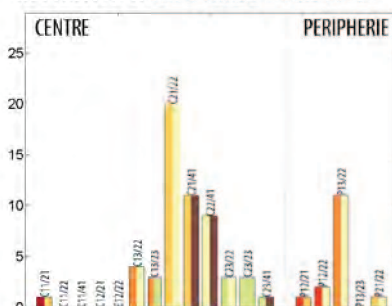
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE MULHOUSE

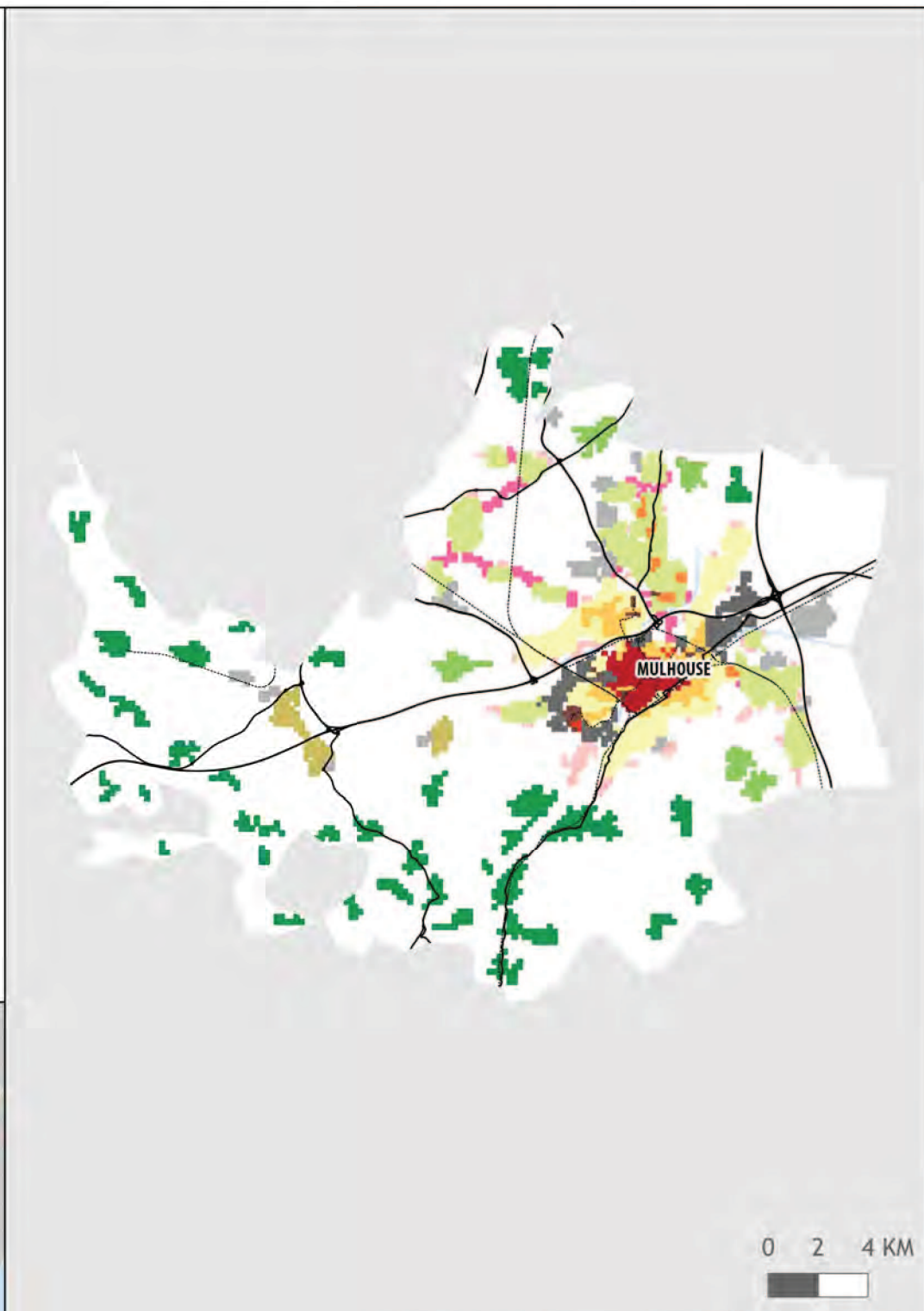
LEGENDE

MORPHOTYPES

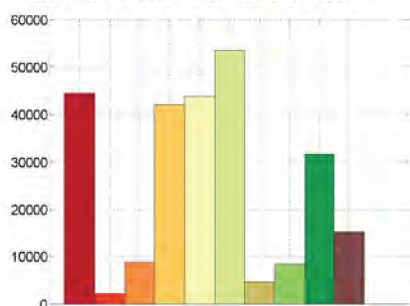
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

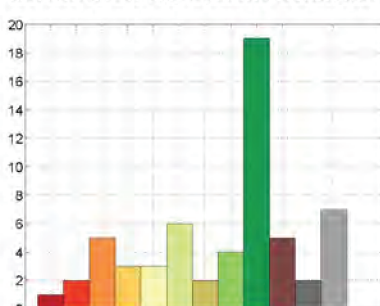
- Voie ferrée
- Route principale
- Surface en eau



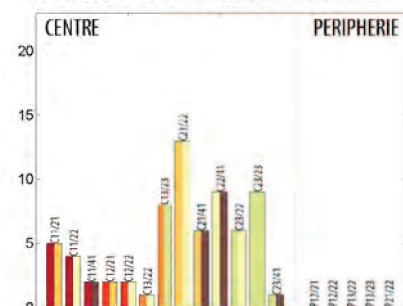
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE NANCY

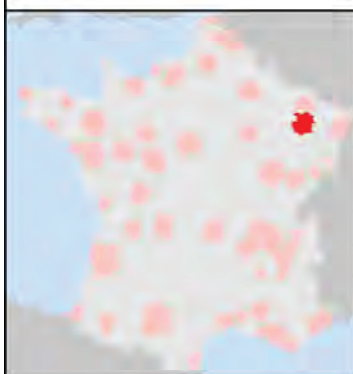
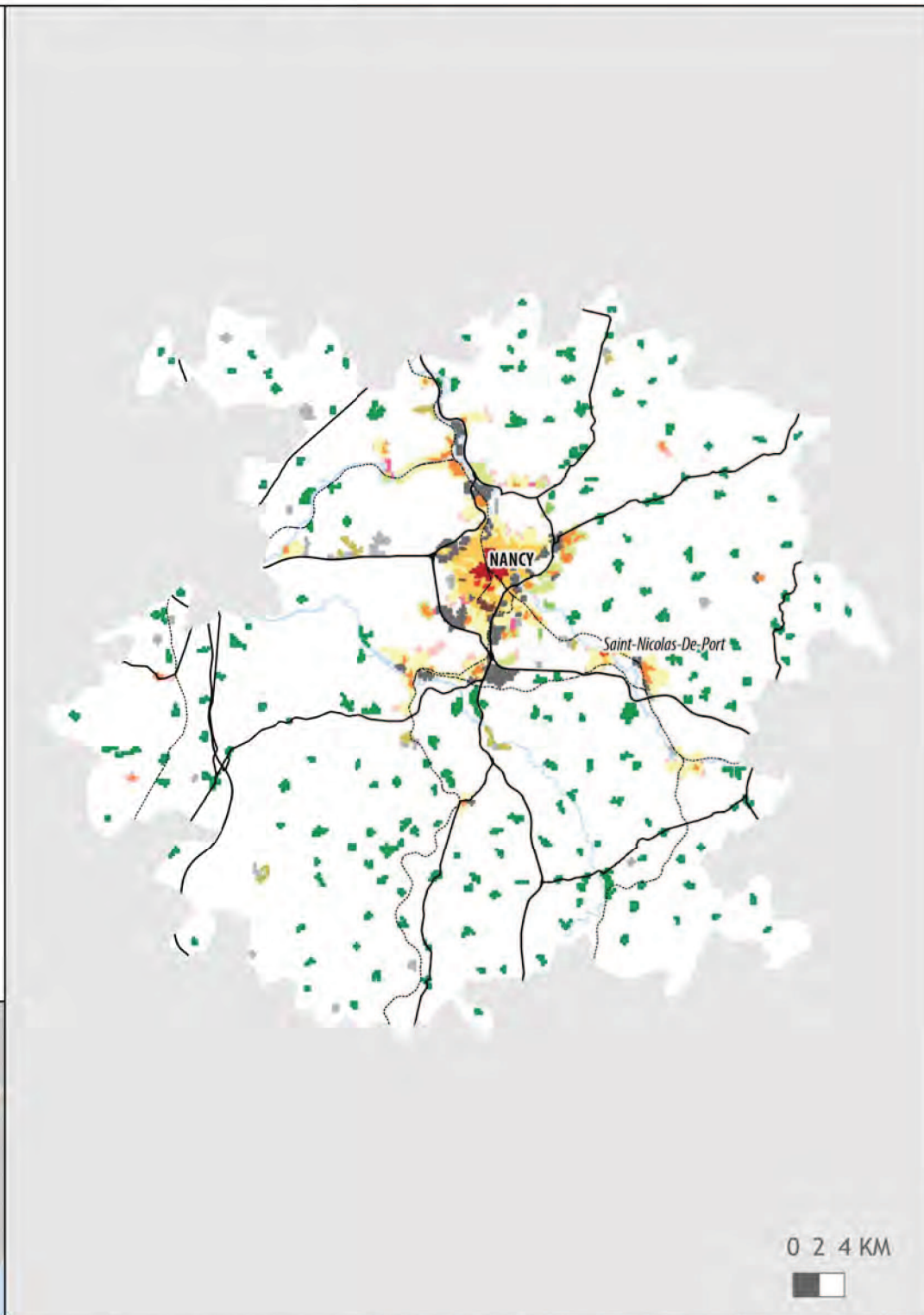
LEGENDE

MORPHOTYPES

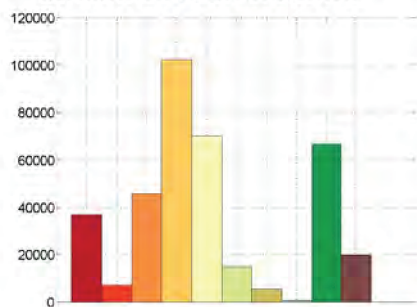
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

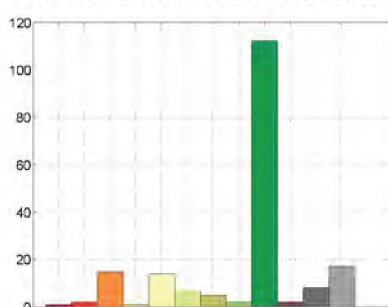
- Voie ferrée
- Route principale
- Surface en eau



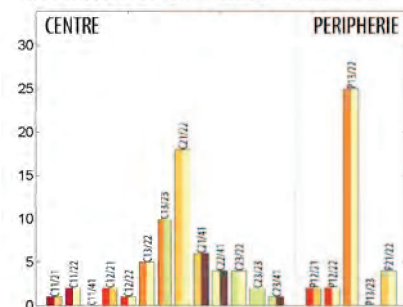
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE NANTES

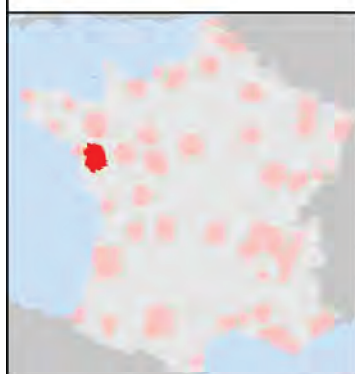
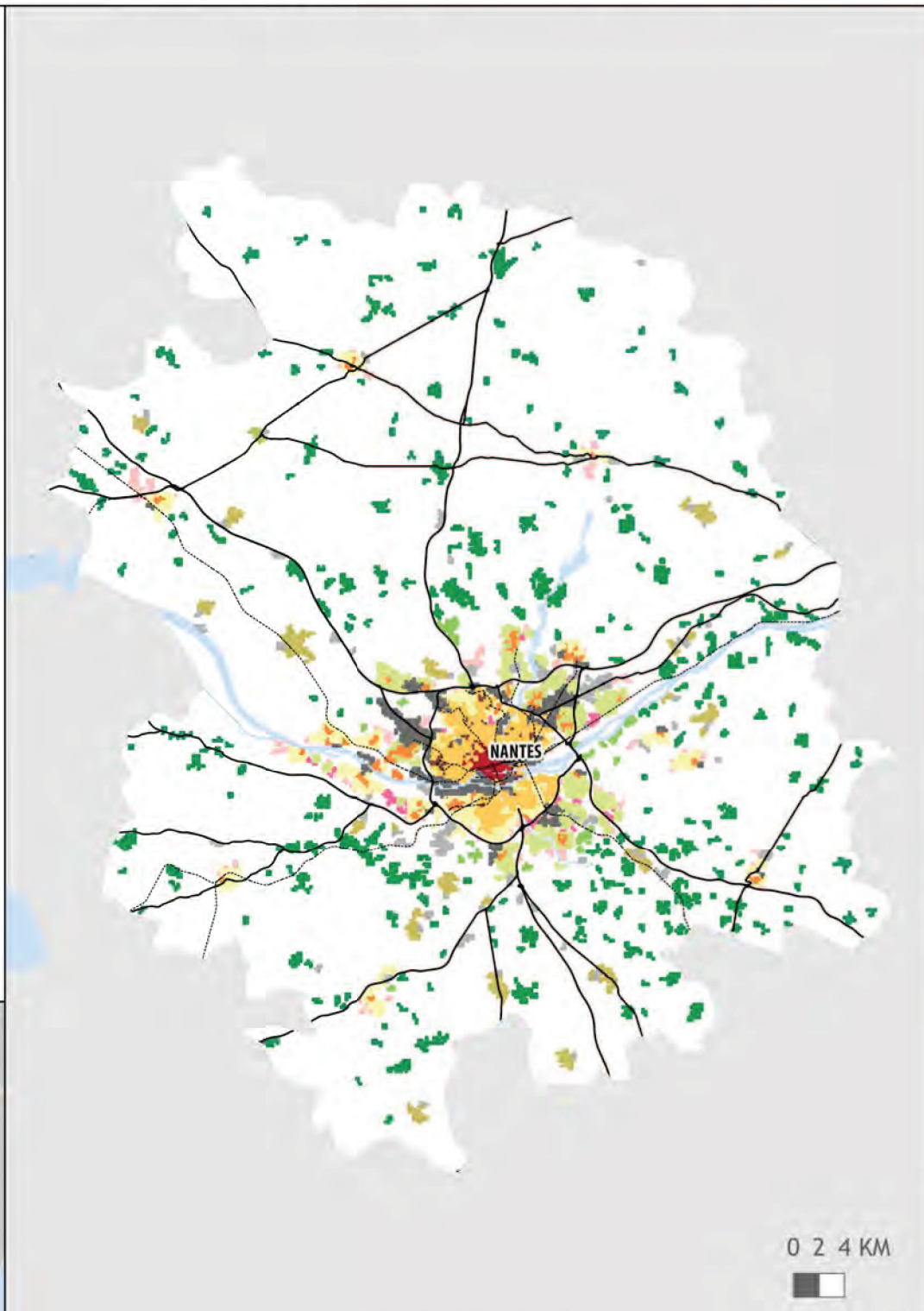
LEGENDE

MORPHOTYPES

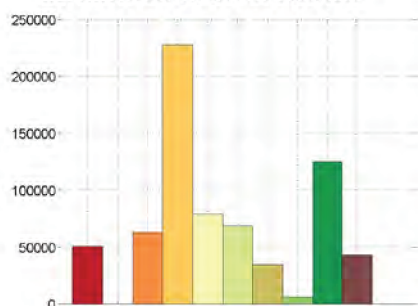
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

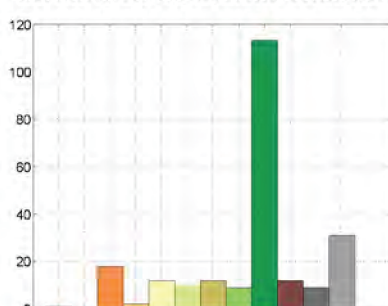
- Voie ferrée
- Route principale
- Surface en eau



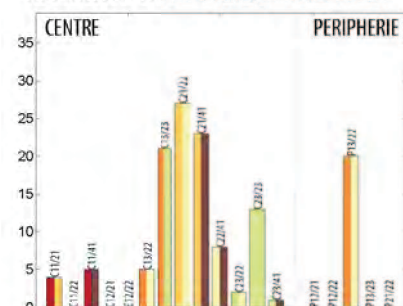
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE NICE

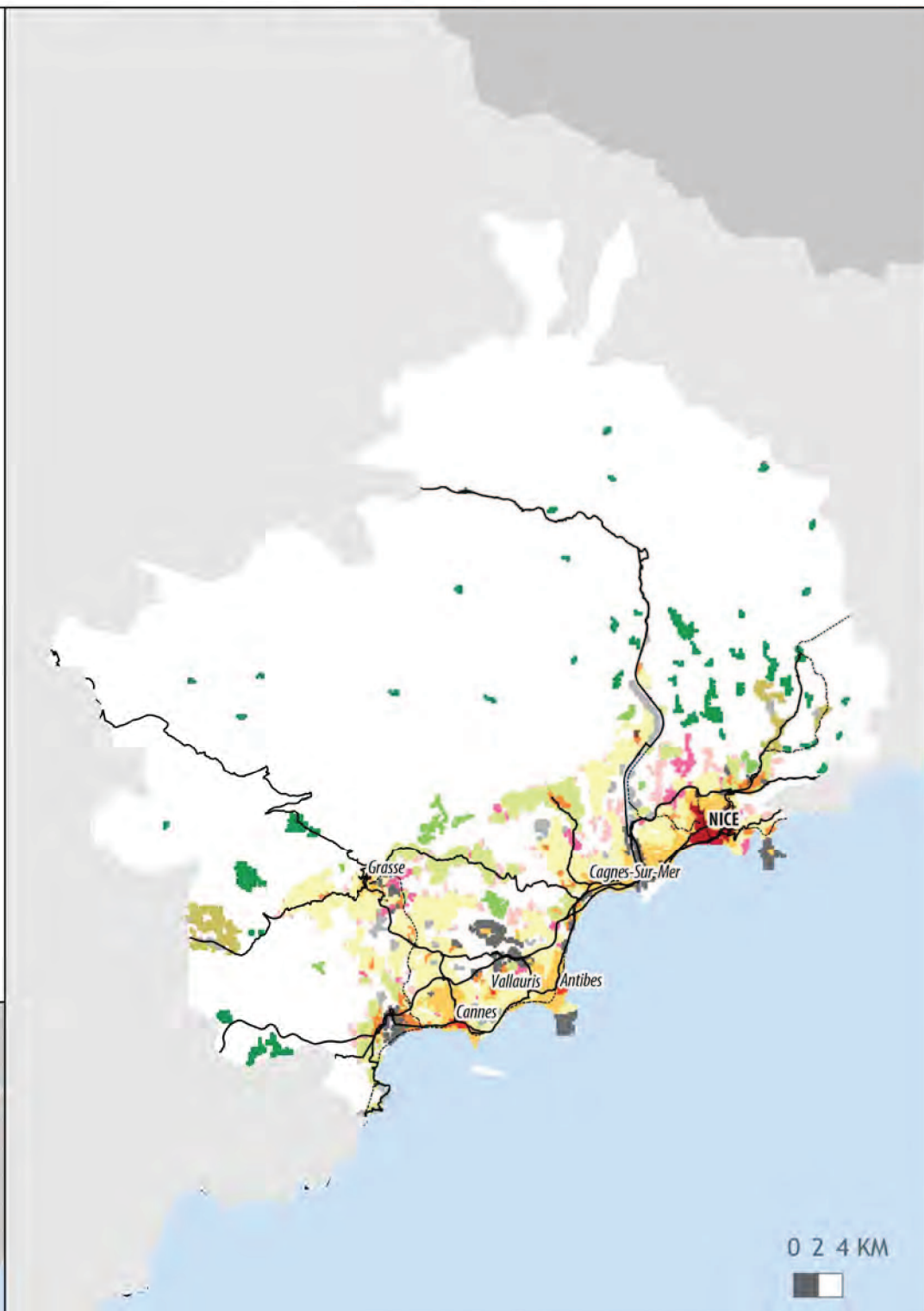
LEGENDE

MORPHOTYPES

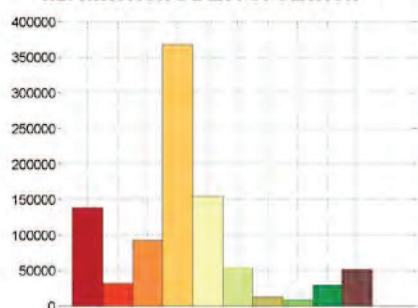
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

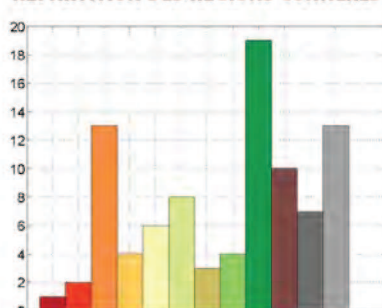
- Voie ferrée
- Route principale
- Surface en eau



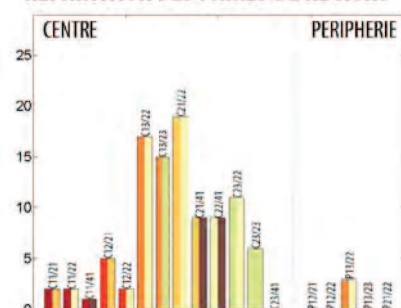
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE NÎMES

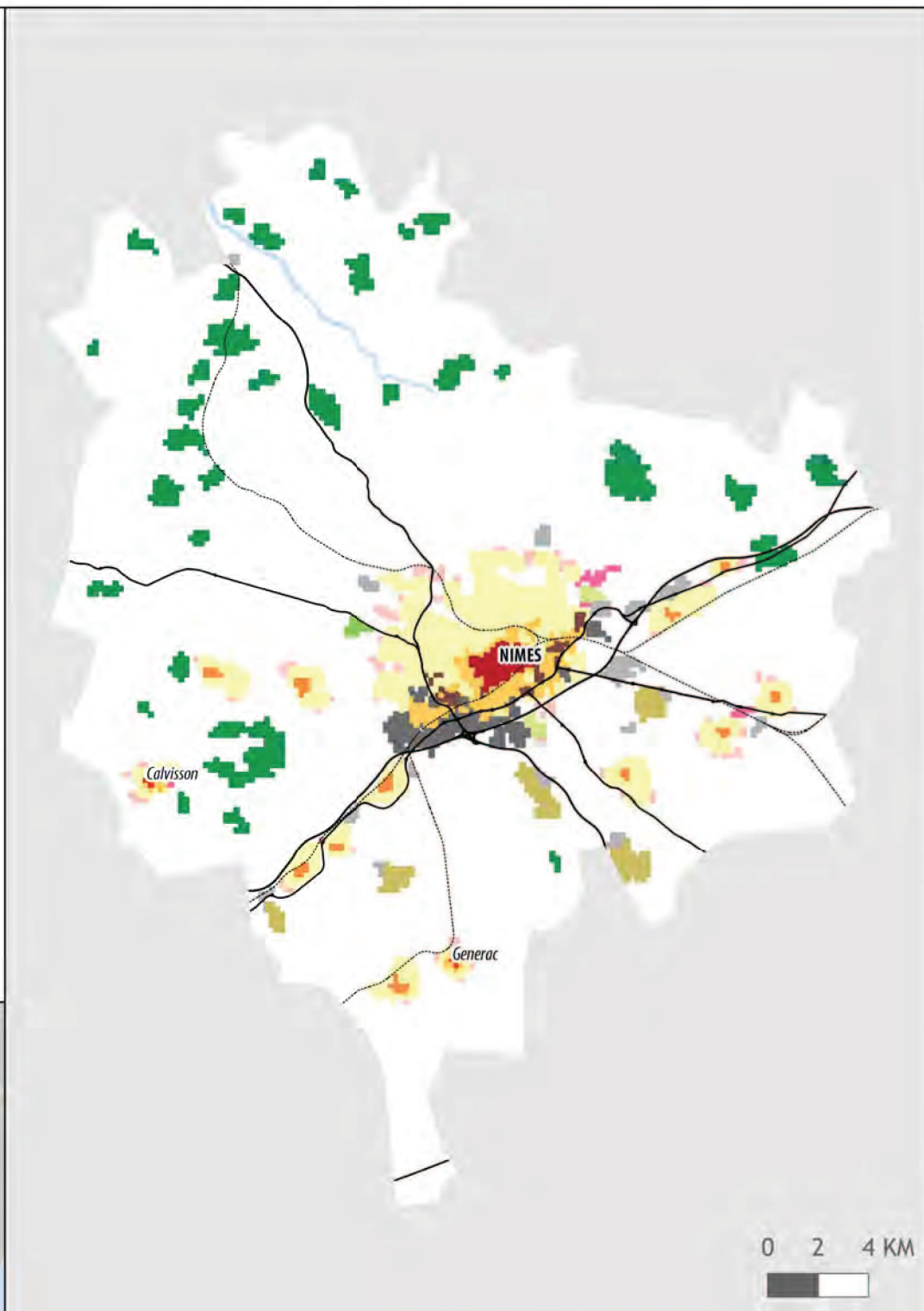
LEGENDE

MORPHOTYPES

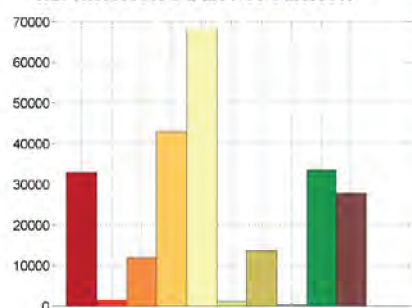
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

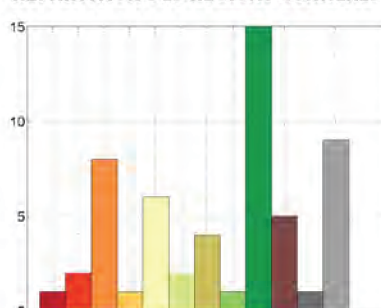
- Voie ferrée
- Route principale
- Surface en eau



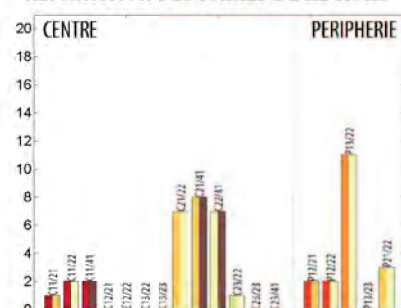
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE D'ORLÉANS

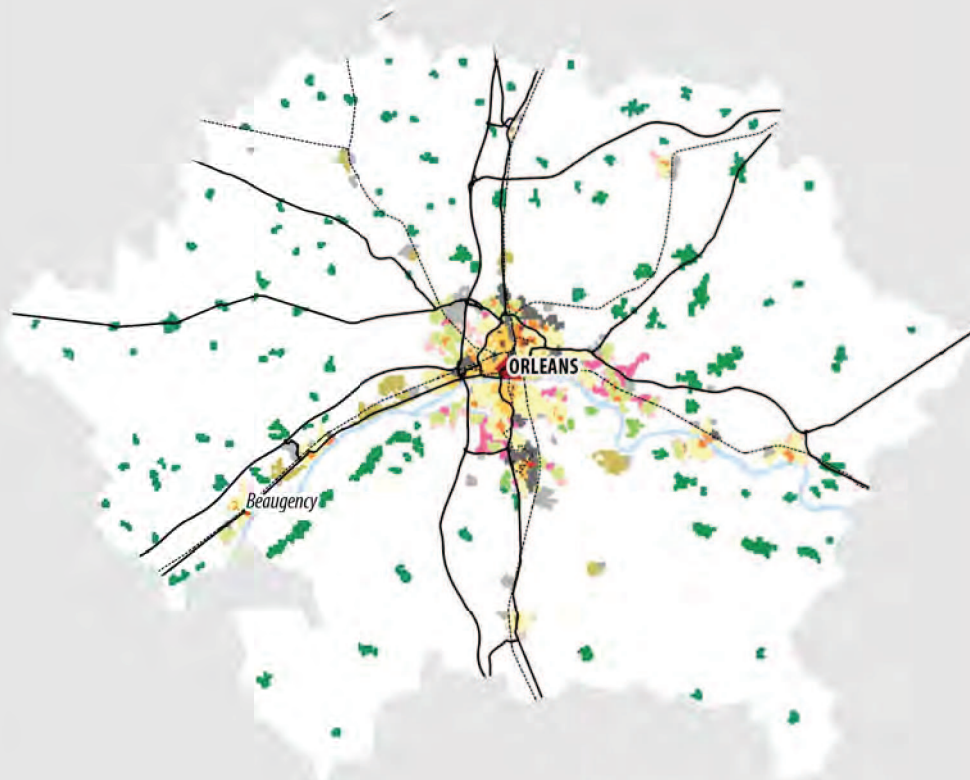
LEGENDE

MORPHOTYPES

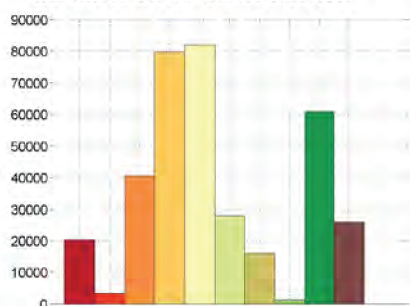
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

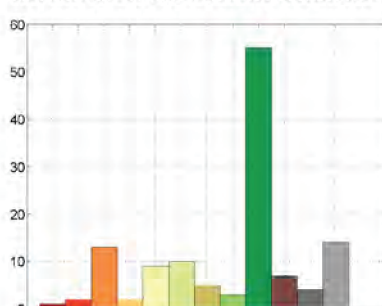
- Voie ferrée
- Route principale
- Surface en eau



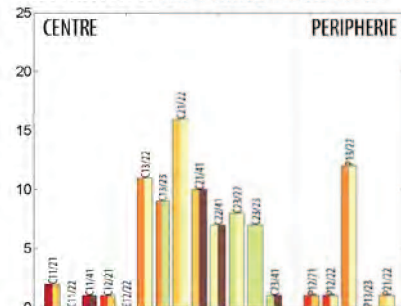
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE PAU

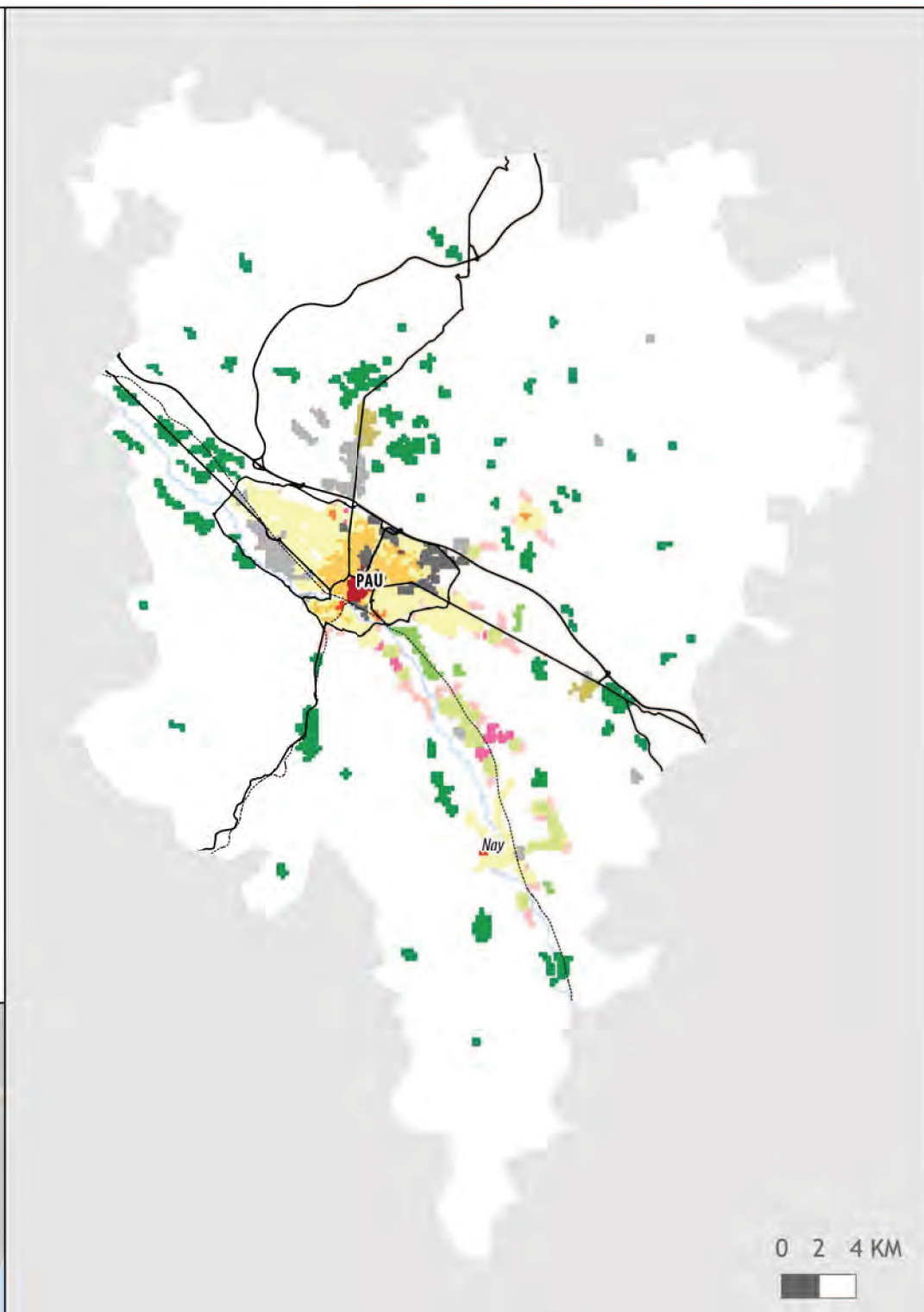
LEGENDE

MORPHOTYPES

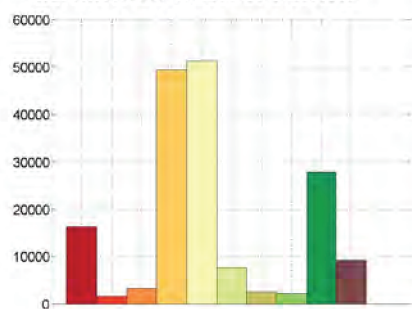
-  Centre principal
-  Centre secondaire
-  Centre tertiaire
-  Couronne dense
-  Couronne peu dense
-  Noyau connecté
-  Grands ensembles
-  Zone d'activité intégrée
-  Zone d'activité périphérique
-  Noyau relais
-  Noyau satellite
-  Noyau isolé
-  Zone connectrice
-  Frange urbaine

HABILLAGE

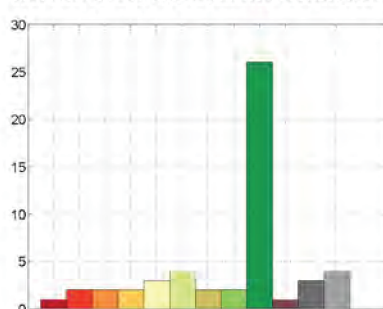
- Voie ferrée
— Route principale
■ Surface en eau



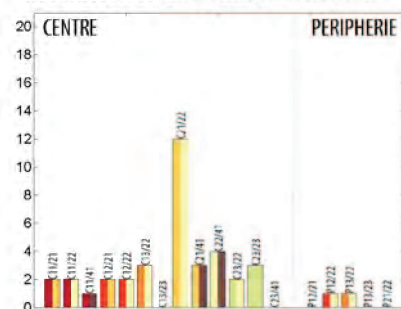
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE PERPIGNAN

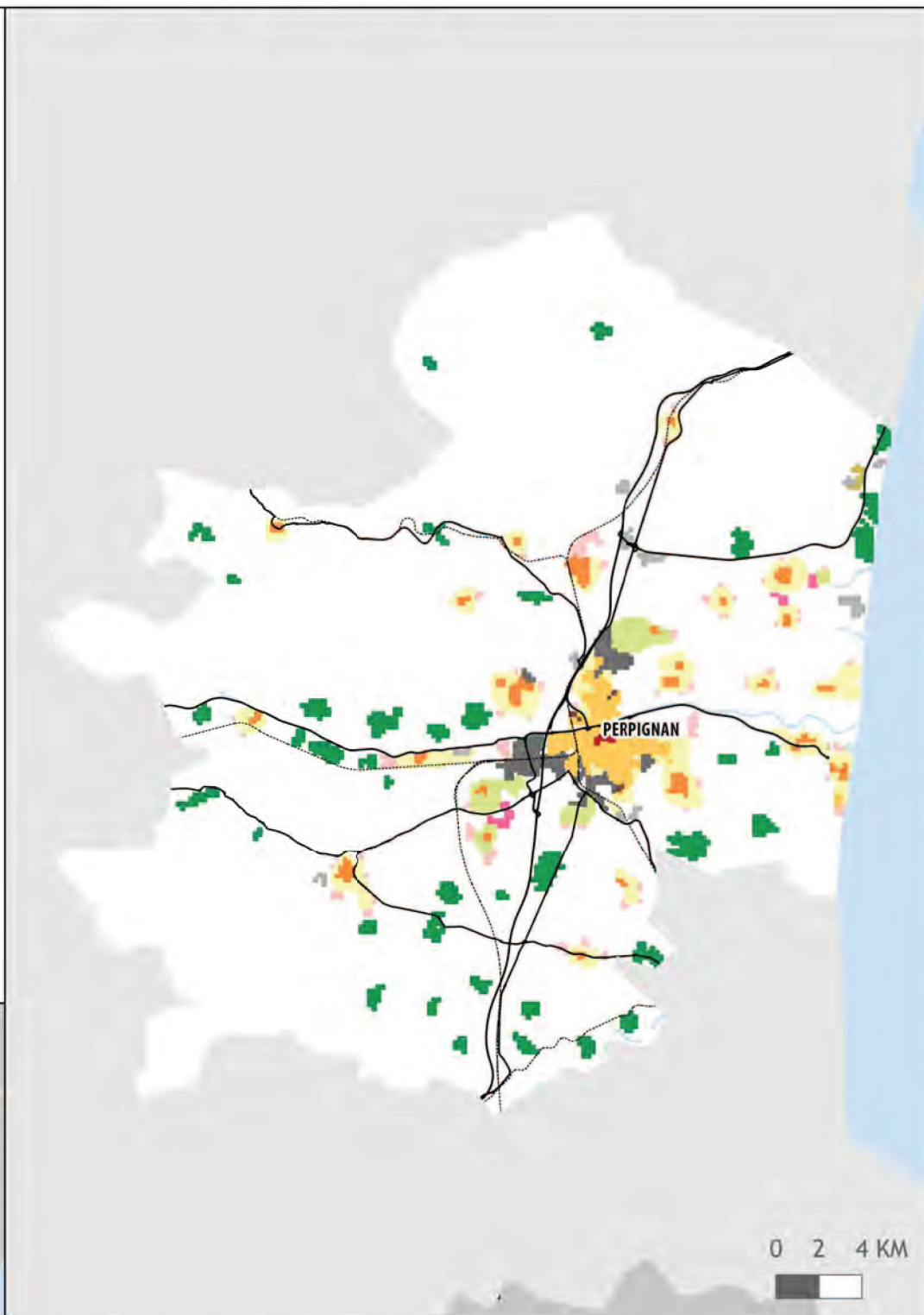
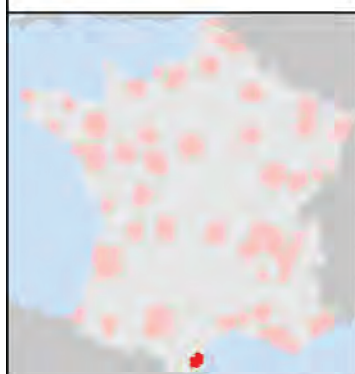
LEGENDE

MORPHOTYPES

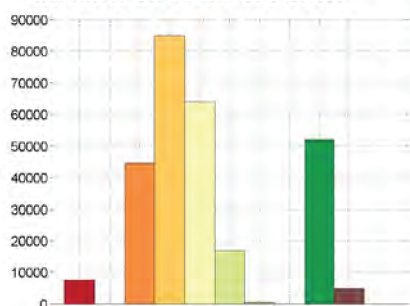
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

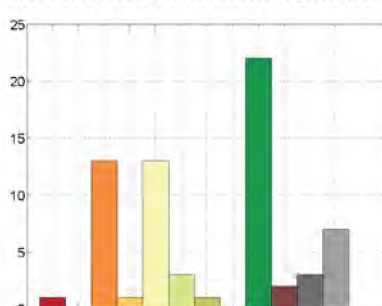
- Voie ferrée
- Route principale
- Surface en eau



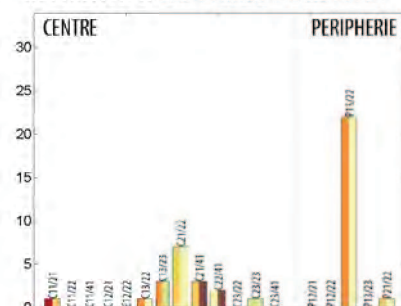
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE POITIERS

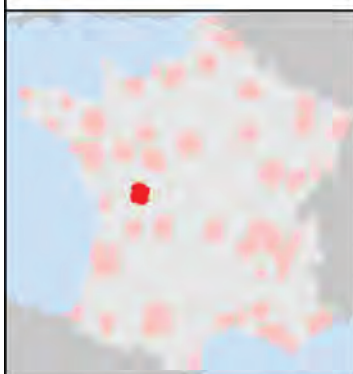
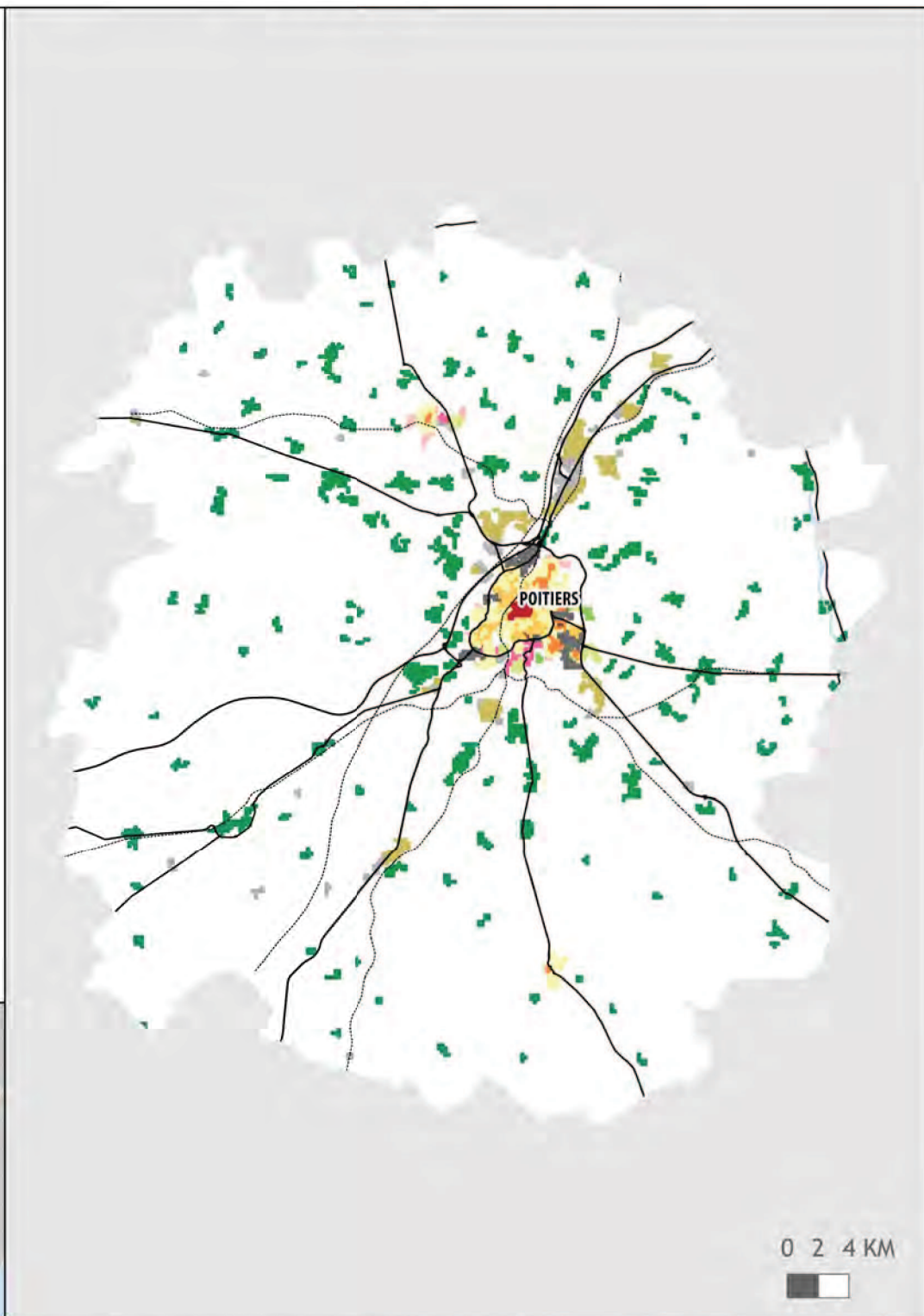
LEGENDE

MORPHOTYPES

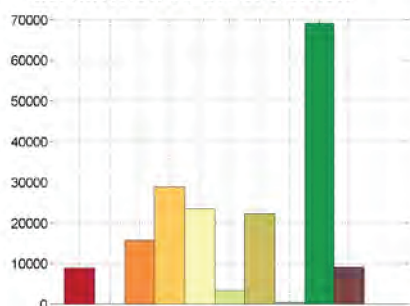
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

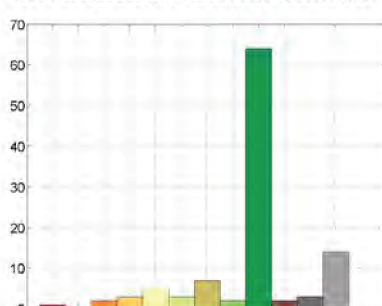
- Voie ferrée
- Route principale
- Surface en eau



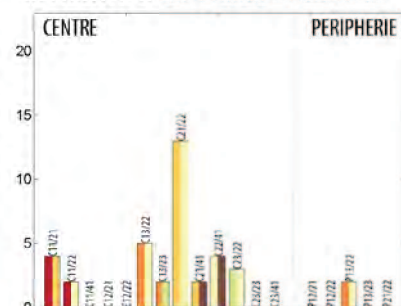
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE REIMS

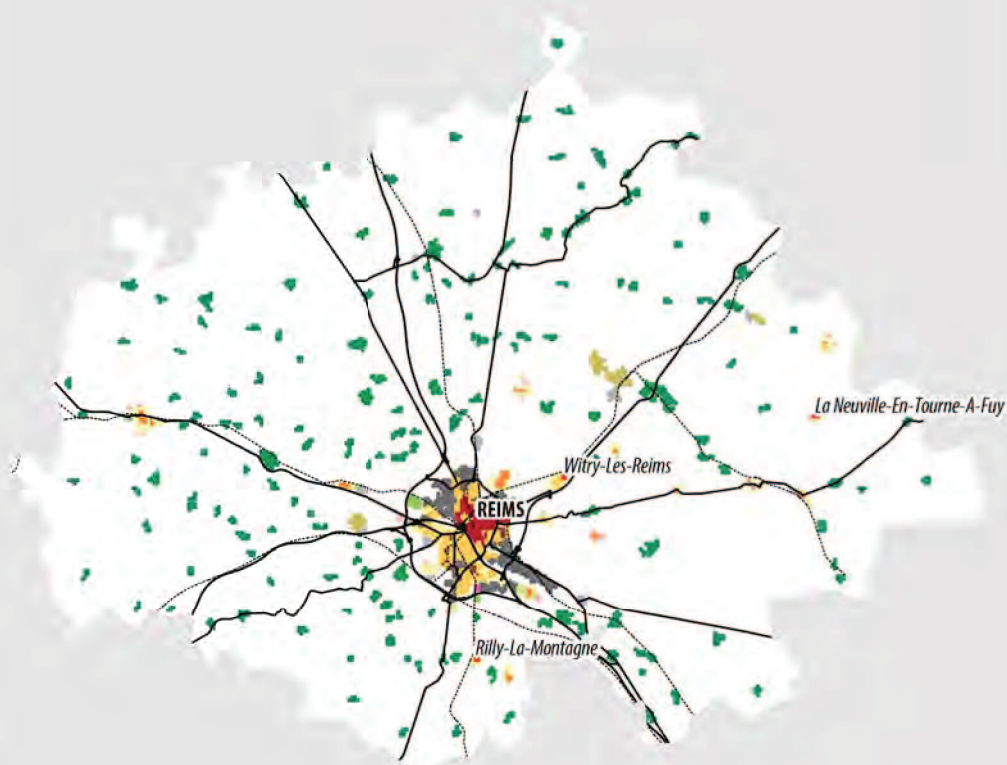
LEGENDE

MORPHOTYPES

- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

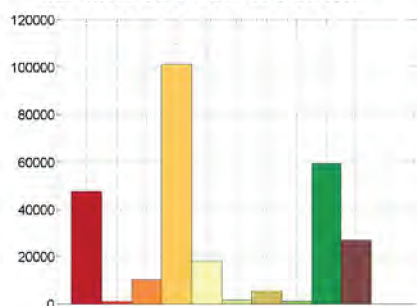
- Voie ferrée
- Route principale
- Surface en eau



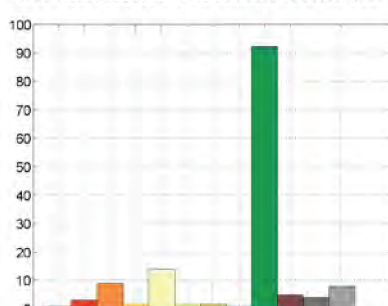
0 2 4 KM



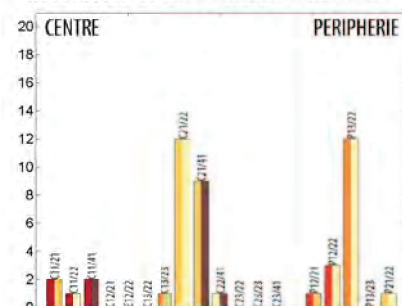
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE RENNES

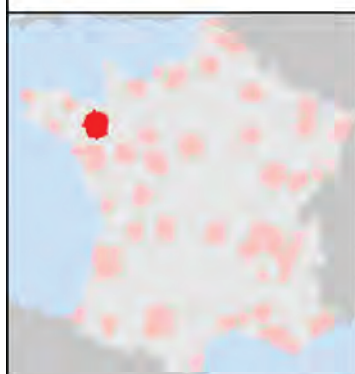
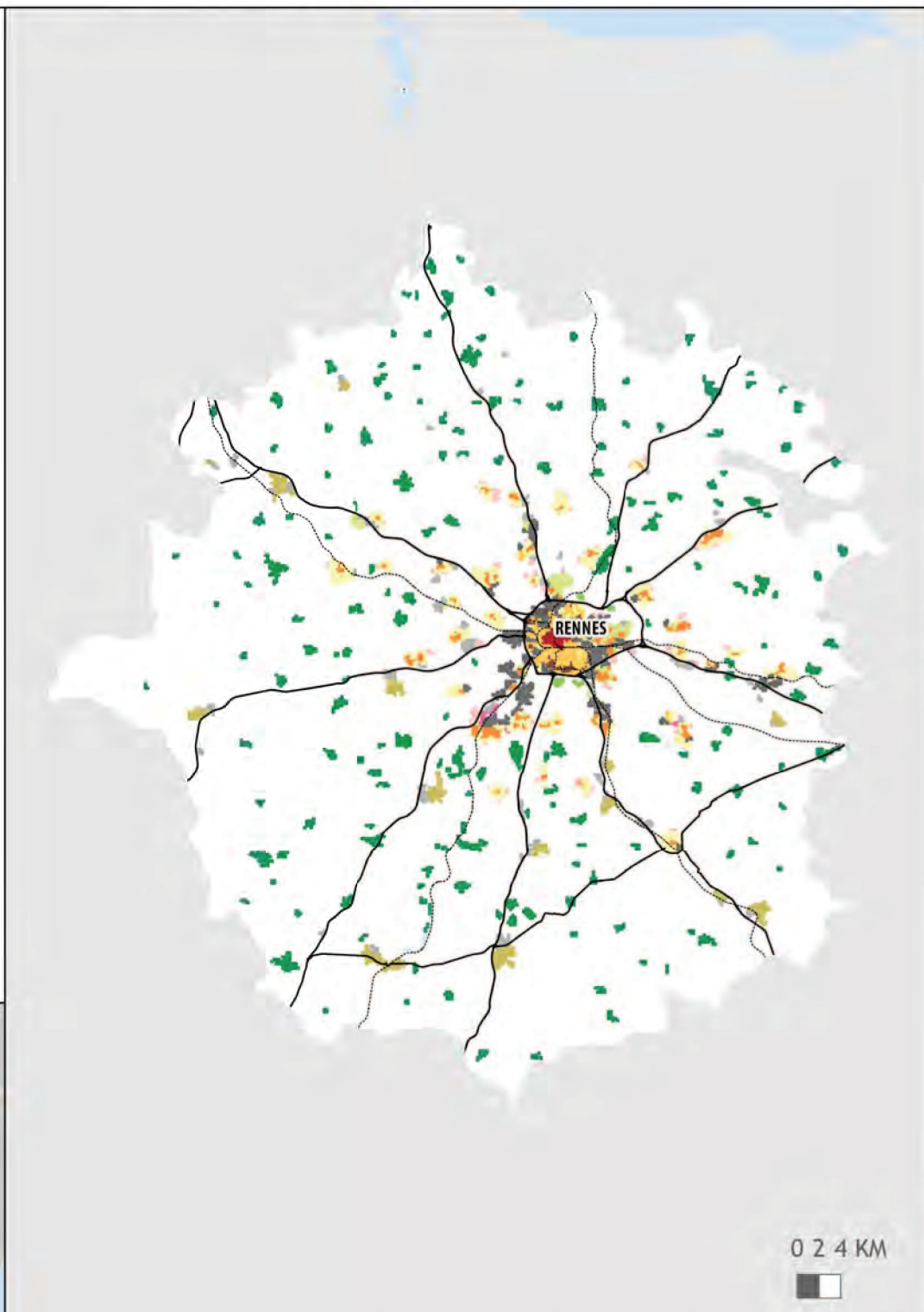
LEGENDE

MORPHOTYPES

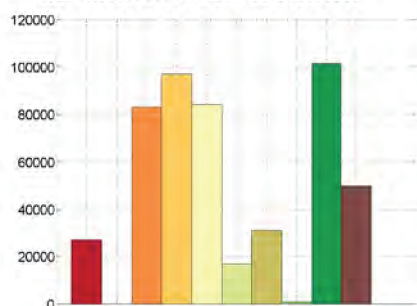
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

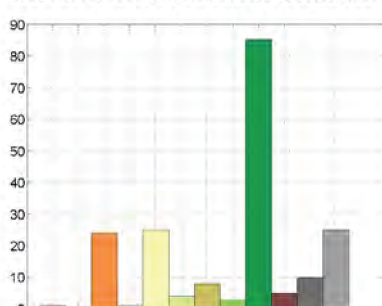
- Voie ferrée
- Route principale
- Surface en eau



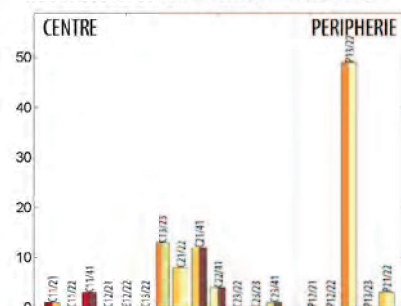
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE ROUEN

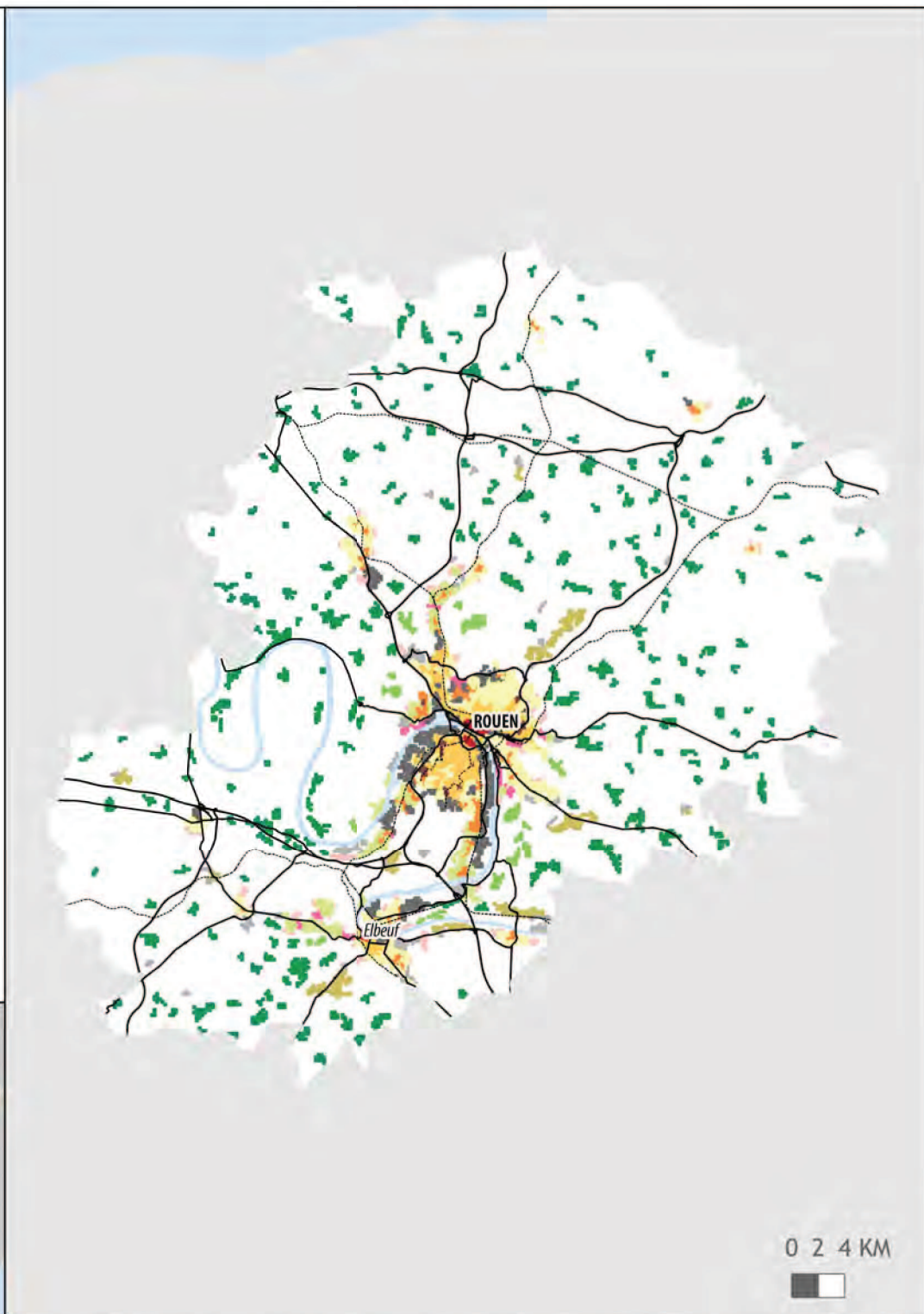
LEGENDE

MORPHOTYPES

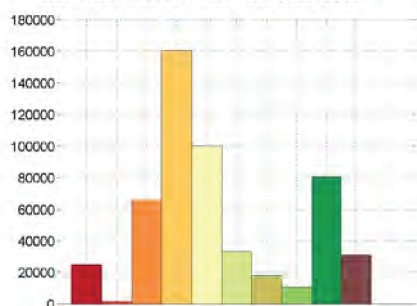
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

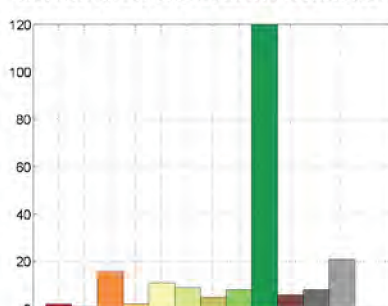
- Voie ferrée
- Route principale
- Surface en eau



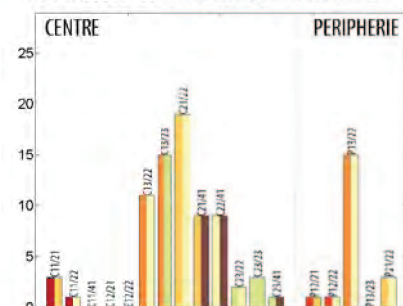
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE SAINT-BRIEUC

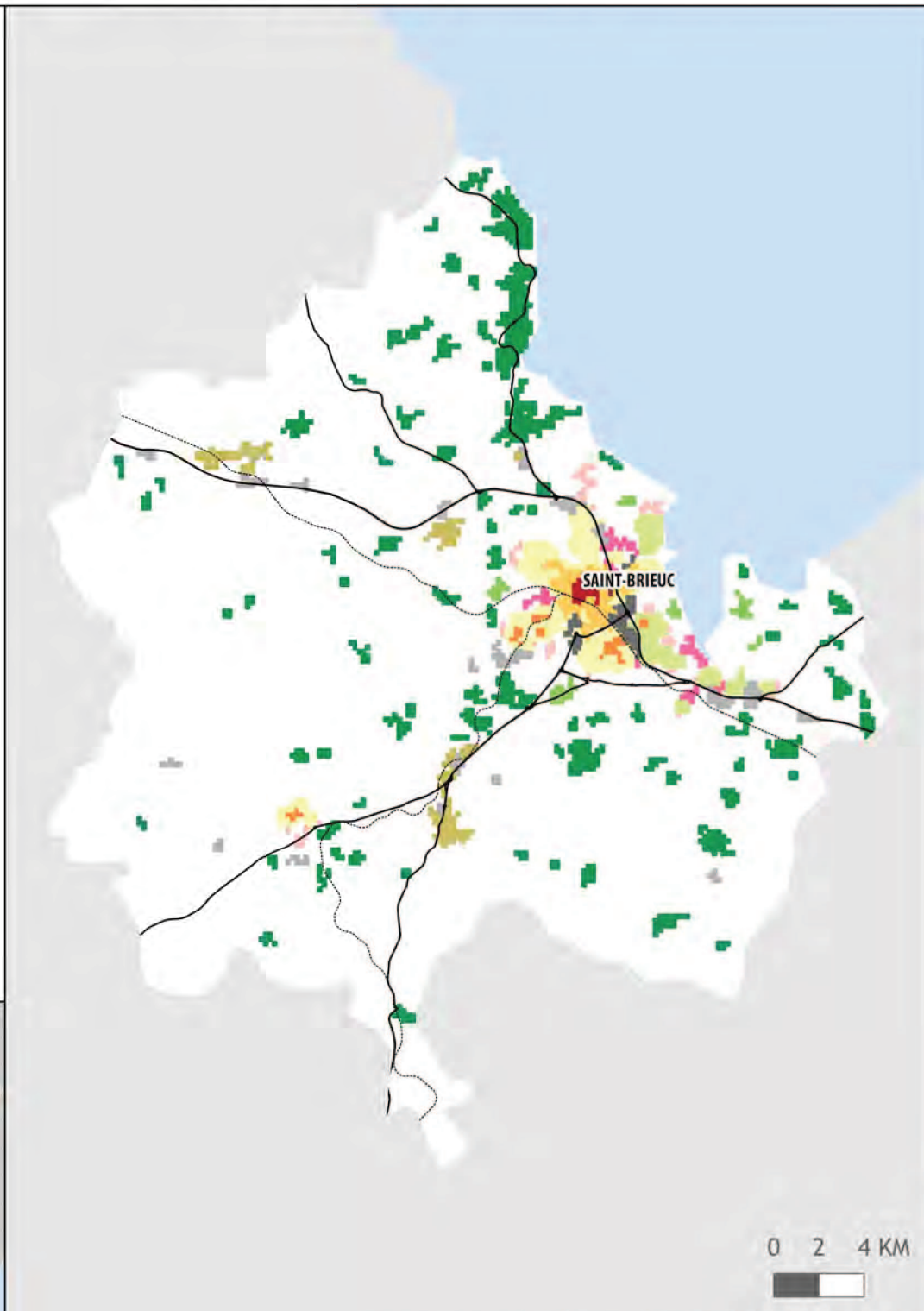
LEGENDE

MORPHOTYPES

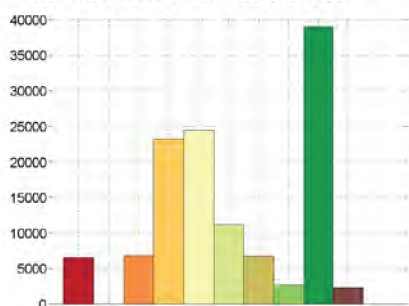
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

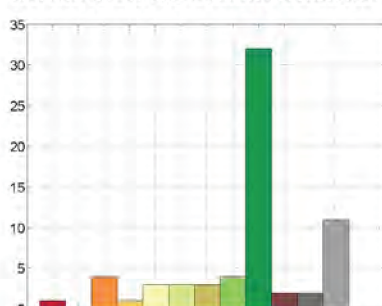
- Voie ferrée
- Route principale
- Surface en eau



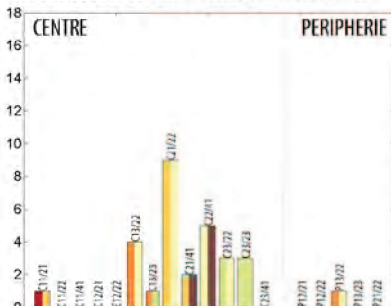
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE SAINT-ÉTIENNE

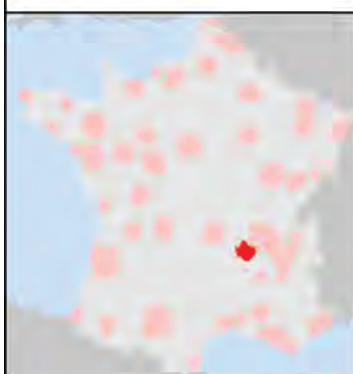
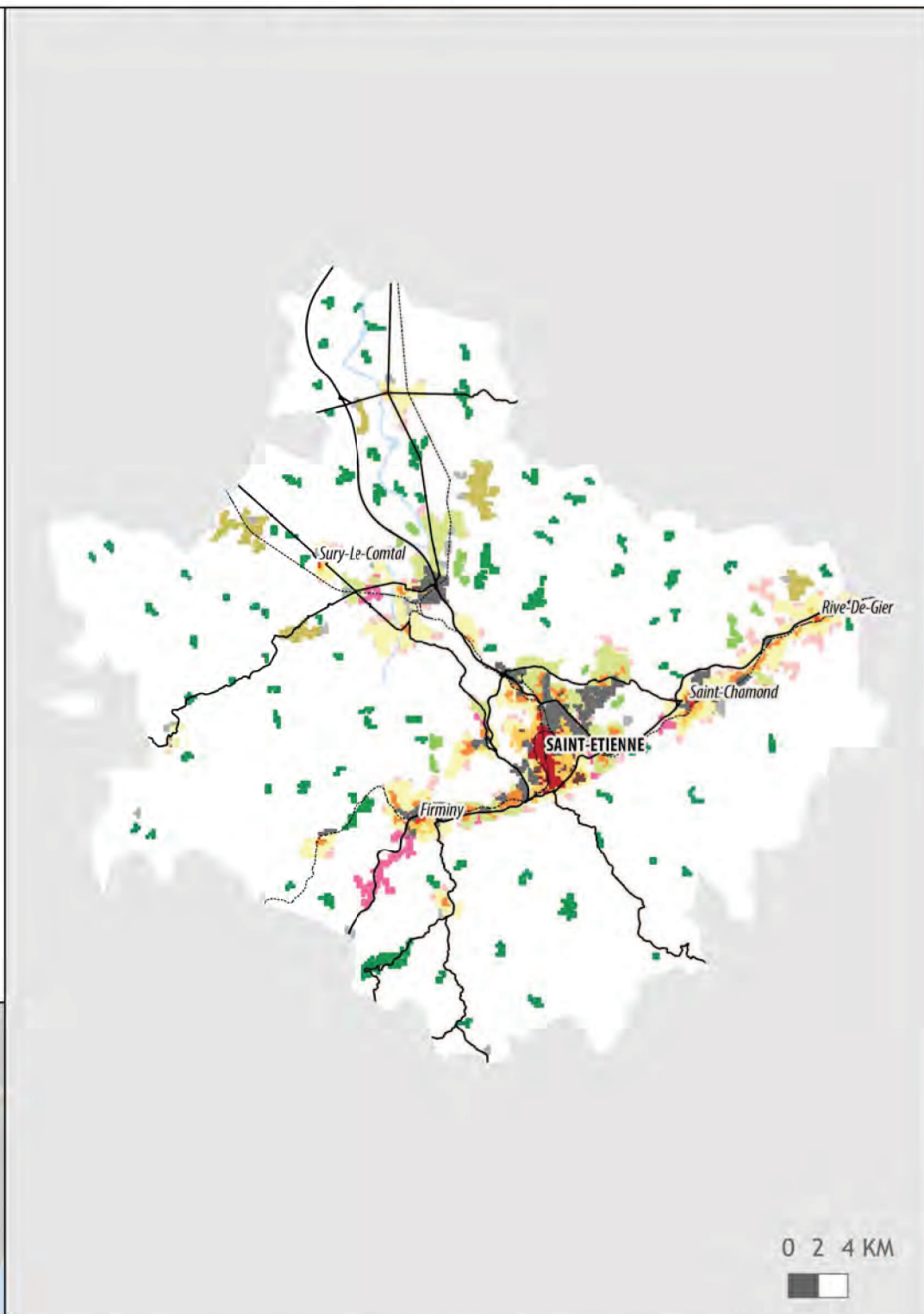
LEGENDE

MORPHOTYPES

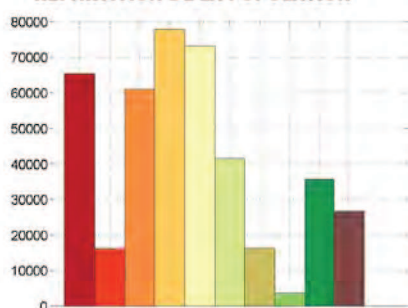
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

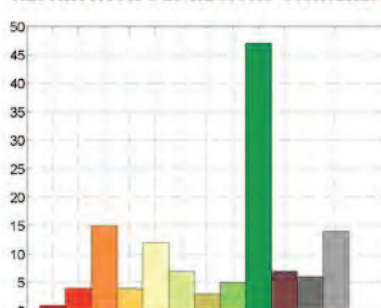
- Voie ferrée
- Route principale
- Surface en eau



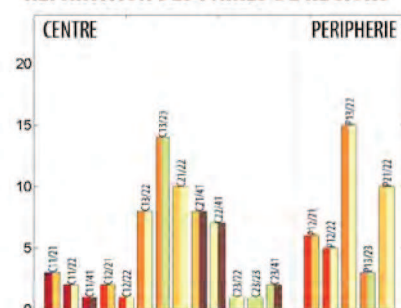
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE SAINT-NAZAIRE

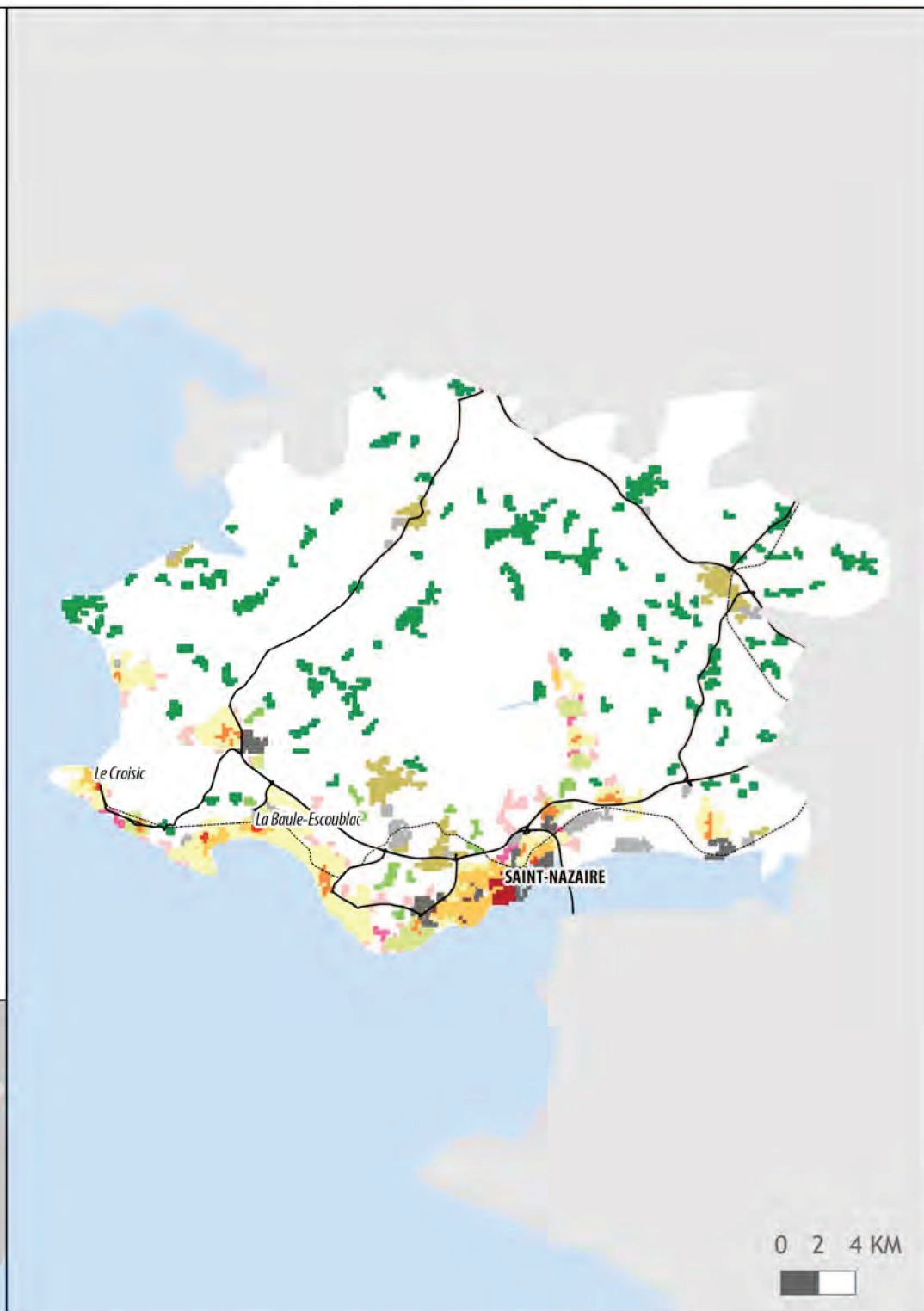
LEGENDE

MORPHOTYPES

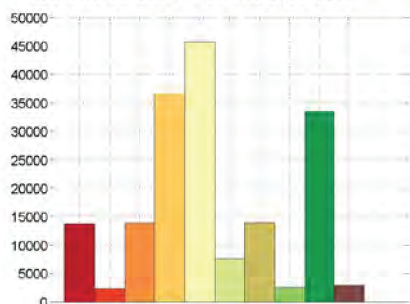
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

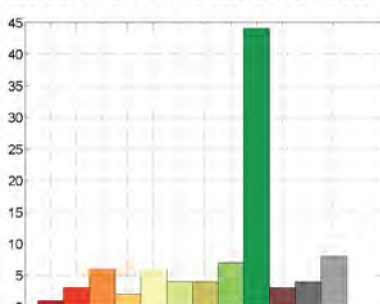
- Voie ferrée
- Route principale
- Surface en eau



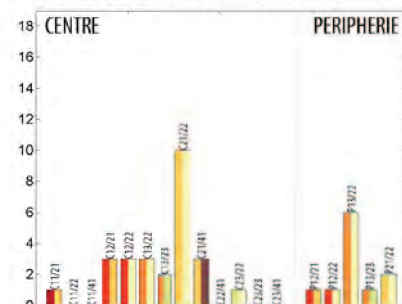
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE STRASBOURG (PARTIE FRANÇAISE)

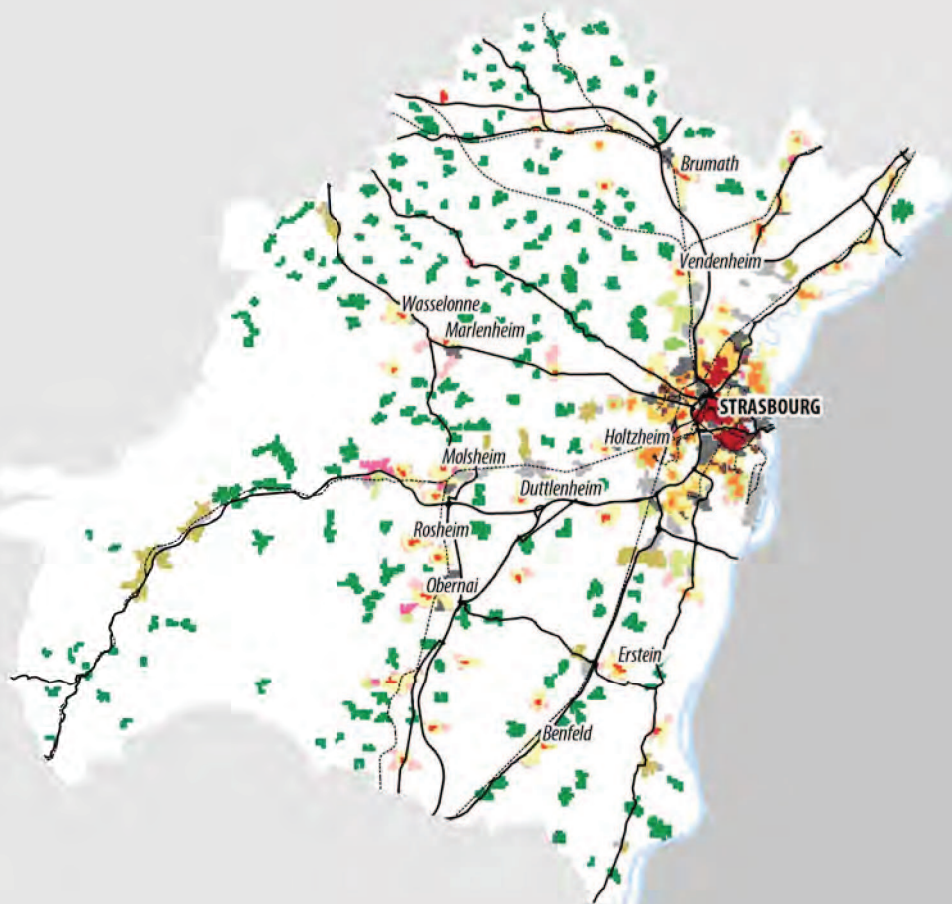
LEGENDE

MORPHOTYPES

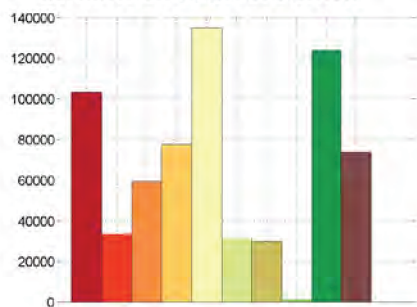
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

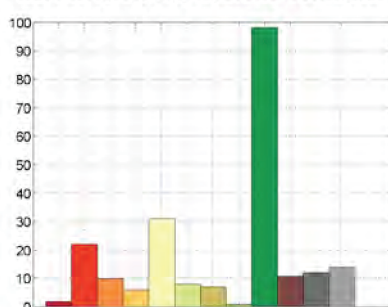
- Voie ferrée
- Route principale
- Surface en eau



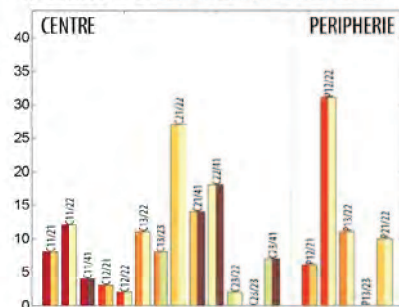
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE TOULON

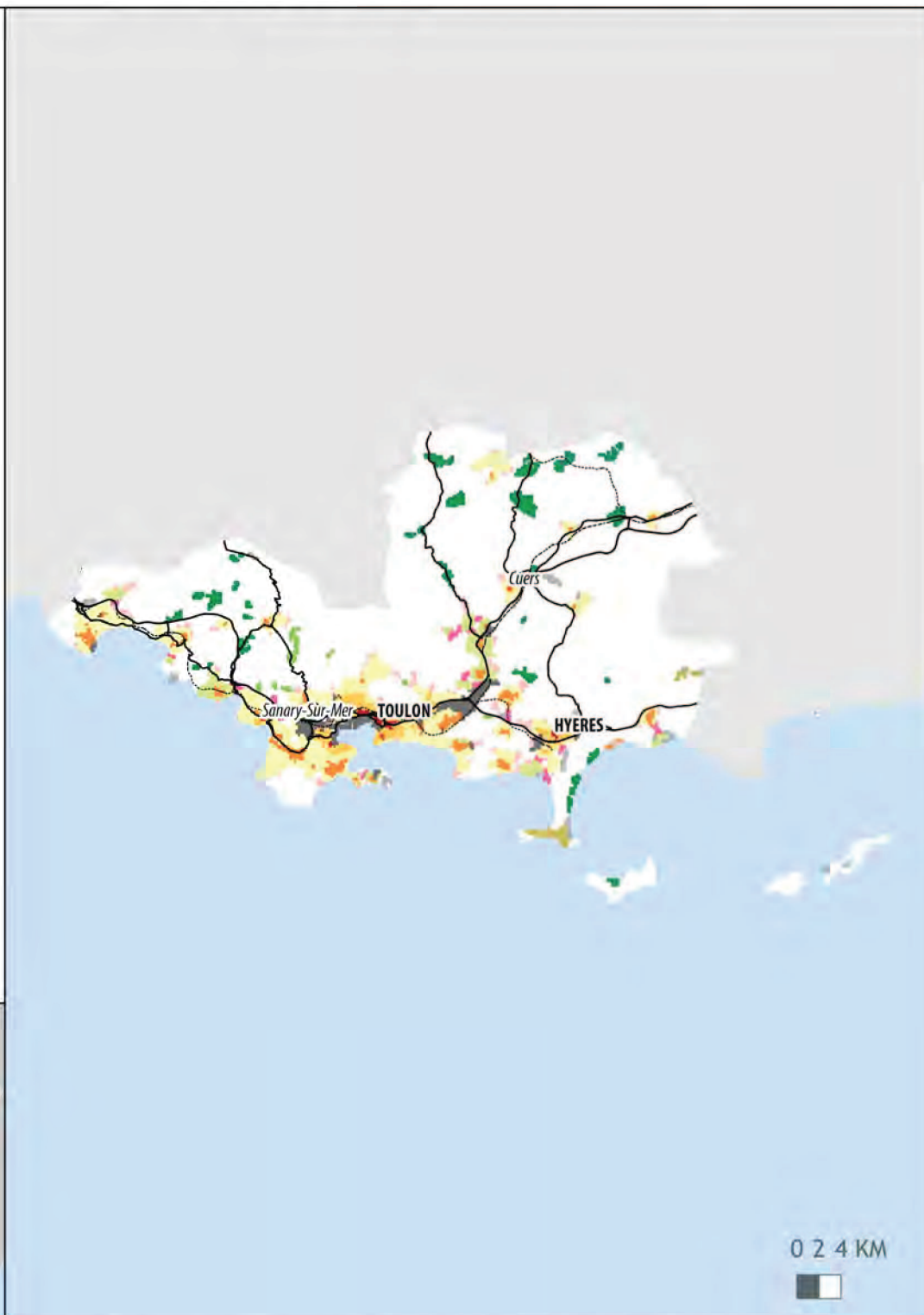
LEGENDE

MORPHOTYPES

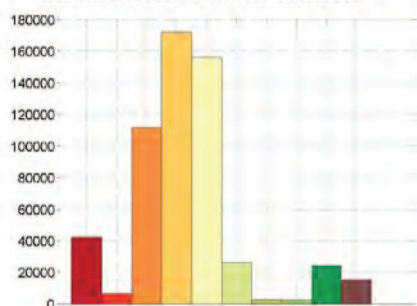
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

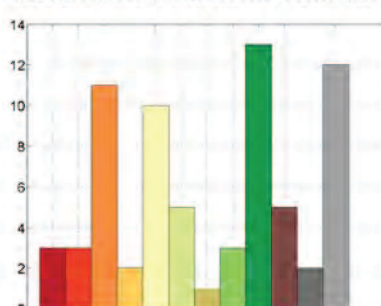
- Voie ferrée
- Route principale
- Surface en eau



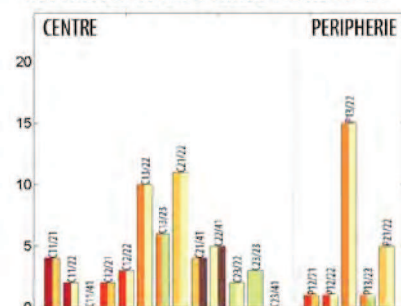
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE TOULOUSE

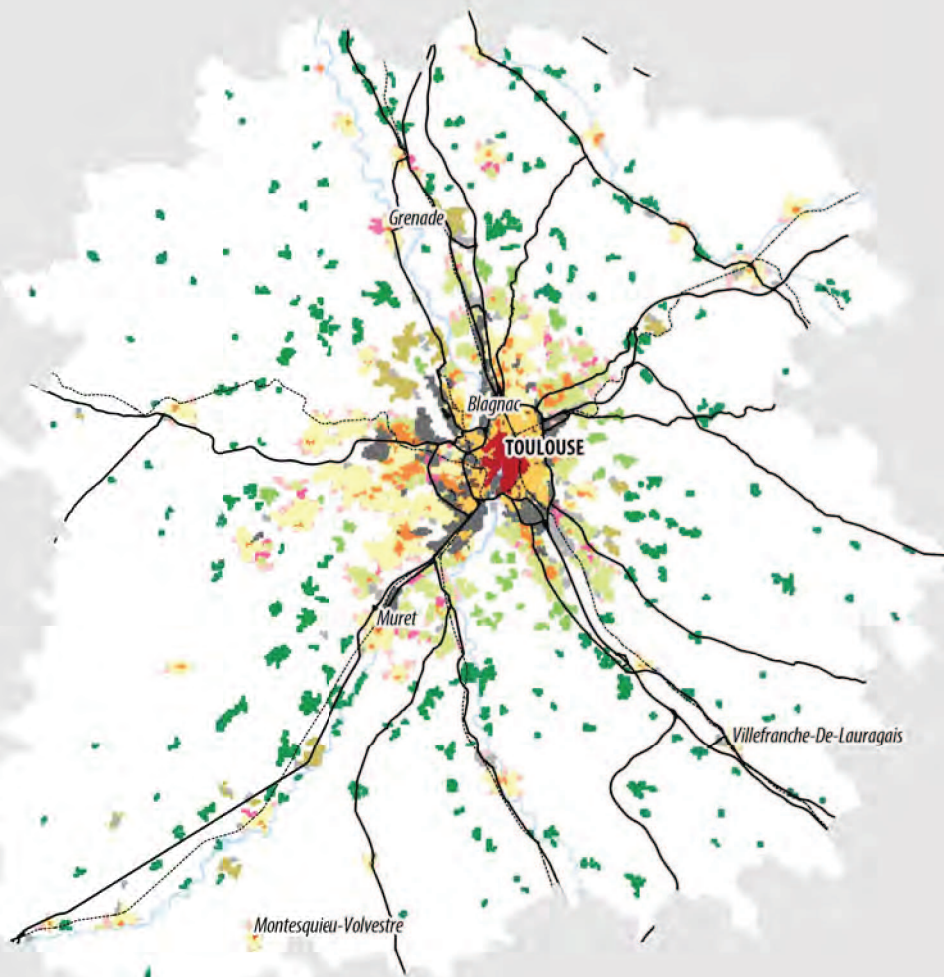
LEGENDE

MORPHOTYPES

- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

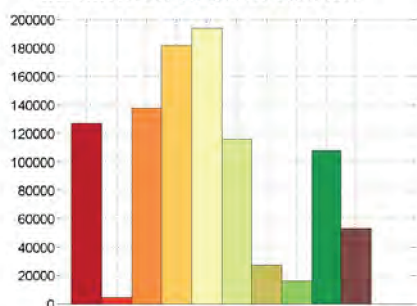
- Voie ferrée
- Route principale
- Surface en eau



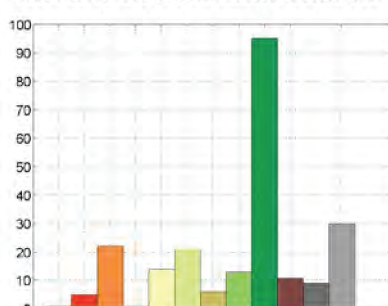
0 2 4 KM



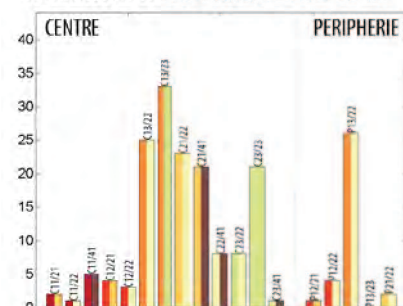
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE TOURS

LEGENDE

MORPHOTYPES

- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

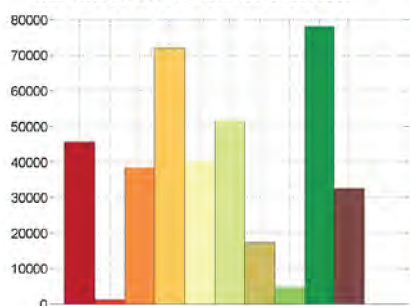
HABILLAGE

- Voie ferrée
- Route principale
- Surface en eau

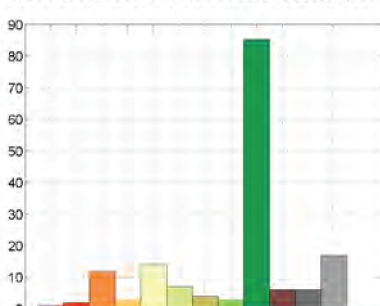


0 2 4 KM

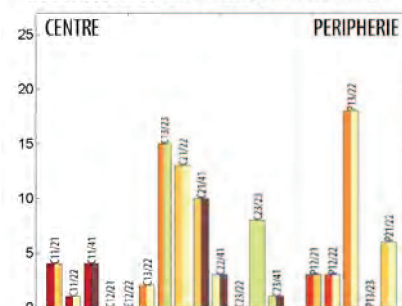
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE TROYES

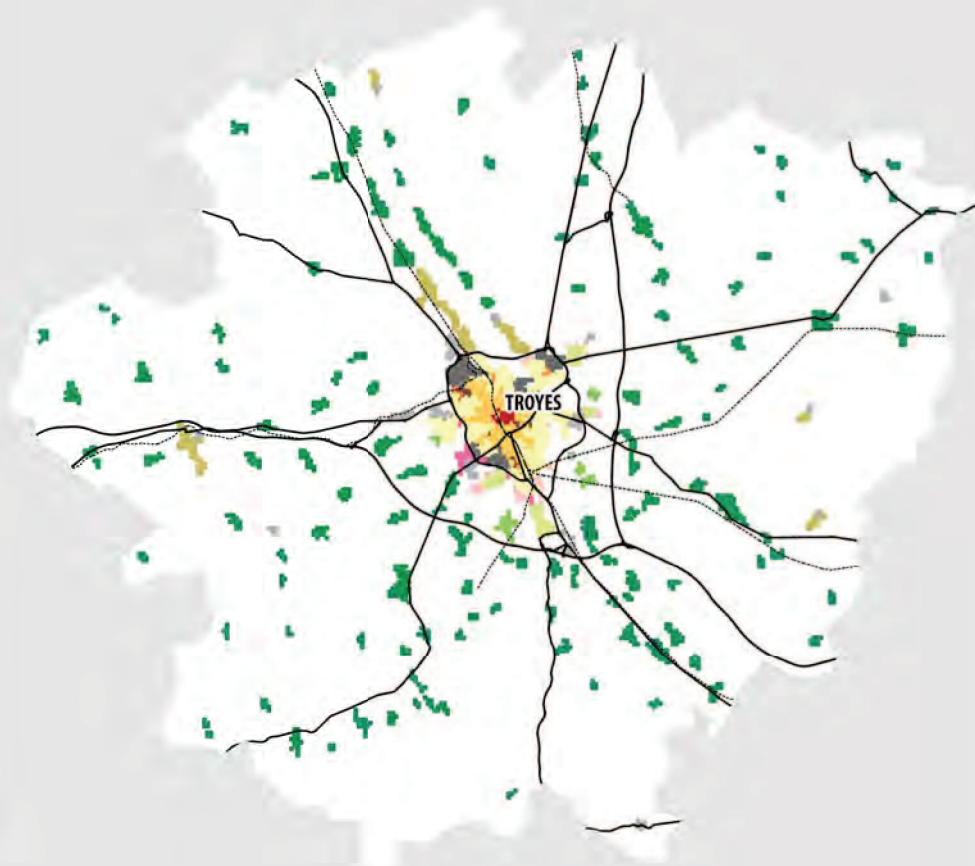
LEGENDE

MORPHOTYPES

- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

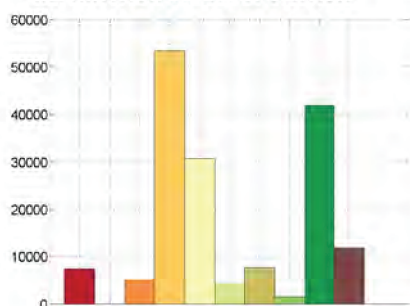
- Voie ferrée
- Route principale
- Surface en eau



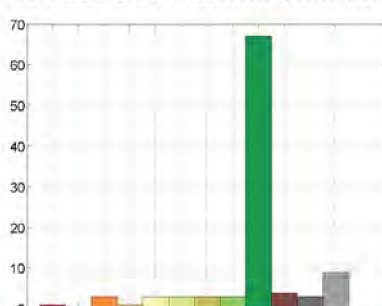
0 2 4 KM



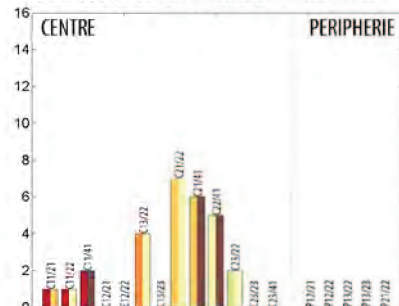
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE VALENCE

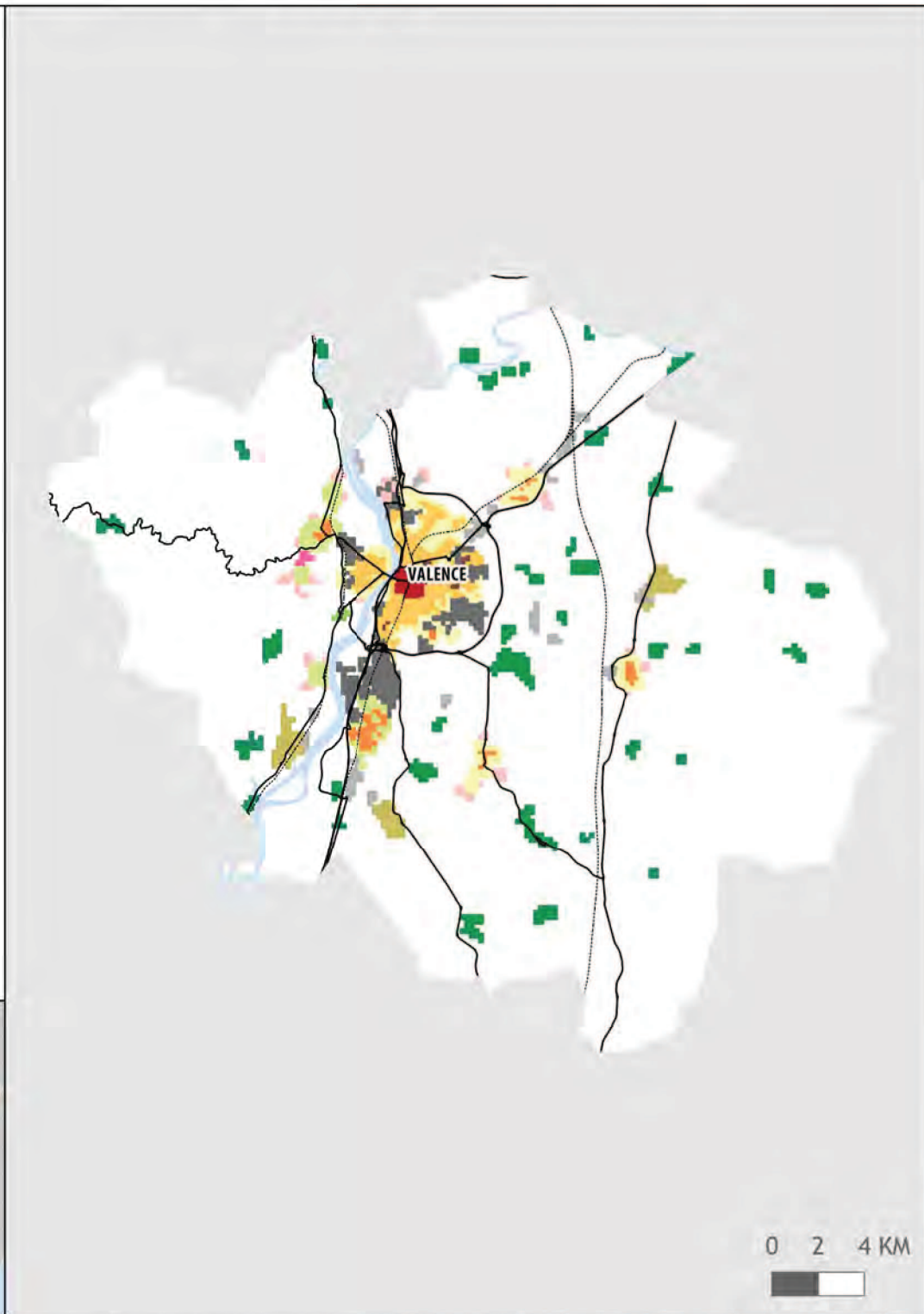
LEGENDE

MORPHOTYPES

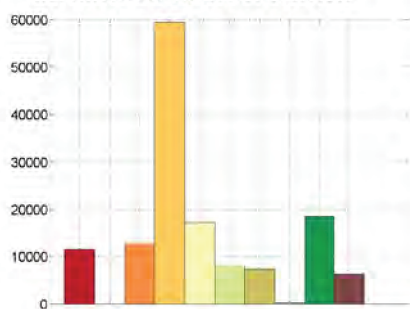
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

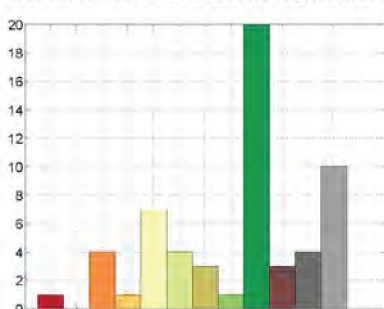
- Voie ferrée
- Route principale
- Surface en eau



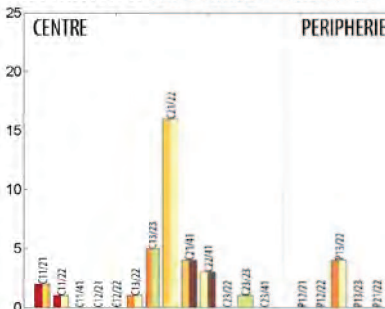
REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS



STRUCTURE URBAINE EN DENSITE DE POPULATION ET DE BATIMENTS

AIRE URBAINE DE VALENCIENNES (PARTIE FRANÇAISE)

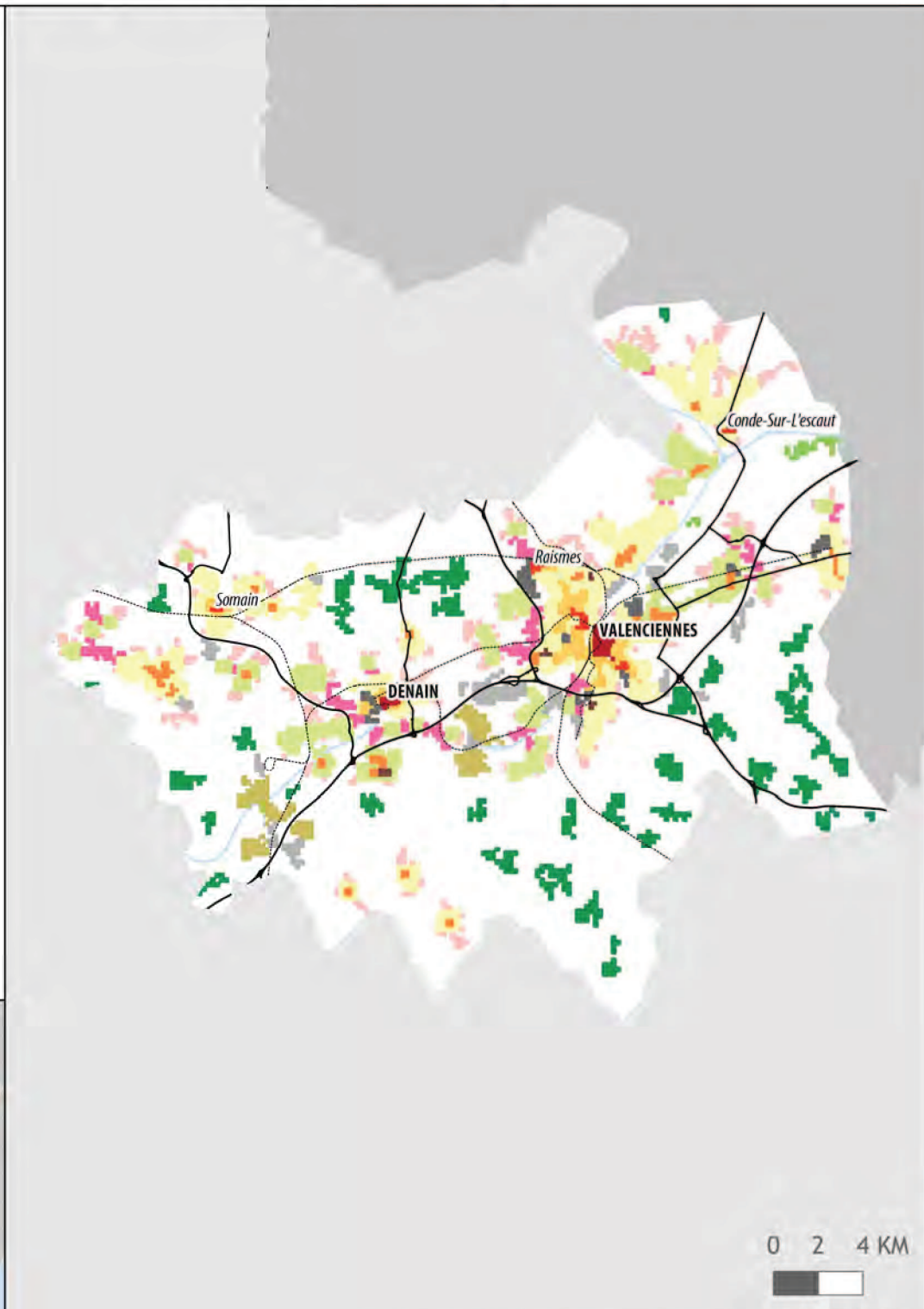
LEGENDE

MORPHOTYPES

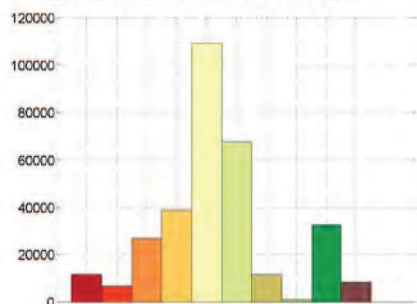
- Centre principal
- Centre secondaire
- Centre tertiaire
- Couronne dense
- Couronne peu dense
- Noyau connecté
- Grands ensembles
- Zone d'activité intégrée
- Zone d'activité périphérique
- Noyau relais
- Noyau satellite
- Noyau isolé
- Zone connectrice
- Frange urbaine

HABILLAGE

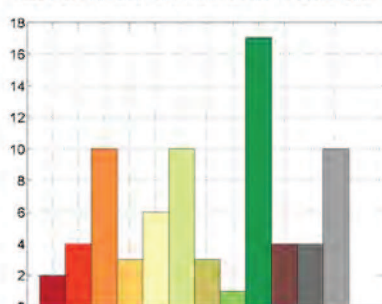
- Voie ferrée
- Route principale
- Surface en eau



REPARTITION DE LA POPULATION



REPARTITION DES REGIONS CONNEXES



REPARTITION DES PAIRES DE REGIONS

